

FINAL REPORT

Interconnections in ULSI: Correlation and Crosstalk

Gary H. Bernstein, Supriyo Bandyopadhyay, and Wolfgang Porod

Department of Electrical Engineering

University of Notre Dame

Notre Dame, IN 46556

Final Report for Period 1 January 1991 - 31 December 1992

ONR Grant Number: N00014-91-J-1505

R&T Number: 4145391 --- 02

164

94-04110



Submitted to

Office of Naval Research, Electronics Division

Arlington, VA 22217

Summary of Research

During the period 1 January 1991 to 31 December 1992, research supported by the Office of Naval Research under Grant Number N00014-91-J-1505 was directed at studying the phenomena of correlation and crosstalk in future ULSI interconnects. As described in the body of this report the coupling phenomena include: (i) line - line coupling, (ii) dot - dot coupling, and (iii) line - dot coupling. We have also addressed the photovoltaic effect, which leads to an induced voltage in a line due to electromagnetic radiation which may be caused by neighboring lines. We also discuss related issues of the fabrication and design of quantum wire and dot structures.

The research supported by the present grant resulted in 36 refereed journal papers and 49 conference publications and presentations. Graduate student support led, in part, to 5 M.S. theses and one Ph.D. dissertation. Individuals whose research was supported, in part, by the present funds include Gary H. Bernstein, S. Bandyopadhyay, Wolfgang Porod, and graduate students Suresh Subramaniam, Nina Telang, Xiaokang Huang, Greg Bazan, and Henry Harbury.

DTIC QUALITY INSPECTED 8

Accession For	
NTIS CRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By <i>Pec Ltr.</i>	
Distribution	
Availability Codes	
Dist	Avail and/or Special
<i>A-1</i>	

Line - Line Coupling

Fabrication of dense interconnect structures

Investigator: G. H. Bernstein

The central mechanism for current drag between metal interconnects is expected to be Coulomb mutual scattering. However, since screening is very strong in metals, in order to observe possible current drag phenomena between metal interconnects, metal patterns must be extremely small and placed very closely together- about 10 nm lines and spaces. Our continued work on high contrast electron resist developers proved critical to the success of achieving the highest possible lithographic resolution.

To this end, we have studied extensively the technology for creating such structures without resorting to such techniques as shadow evaporation. Direct fabrication of such patterns is necessitated by the need to make connections to the individual lines through which currents are sent and measured. Our basic tool is electron beam lithography of poly (methyl methacrylate) (PMMA). The two central issues to creating very dense patterns as described above are proximity effects and the strength of PMMA "walls".

Proximity effects are important since the lines are so closely spaced that electron scattering phenomena enter a new regime not commonly studied for the fabrication of conventional integrated circuits. Most proximity effect studies use a double Gaussian model taking into account short range forward scattering and long range backscattering. Usually not considered, however, are the high energy secondary electrons of intermediate range. We performed a very careful study fitting curves to include a third Gaussian component due to the fast secondary electrons. Figure 1 shows the relative dose necessary to create lines of equal width as a function of the grating pitch. It was found that the use of the third, intermediate range, component gave a much better fit to the data than the use of only two Gaussian components. We believe this is the only study to date of the proximity effect on the size scales of a few tens of nm.

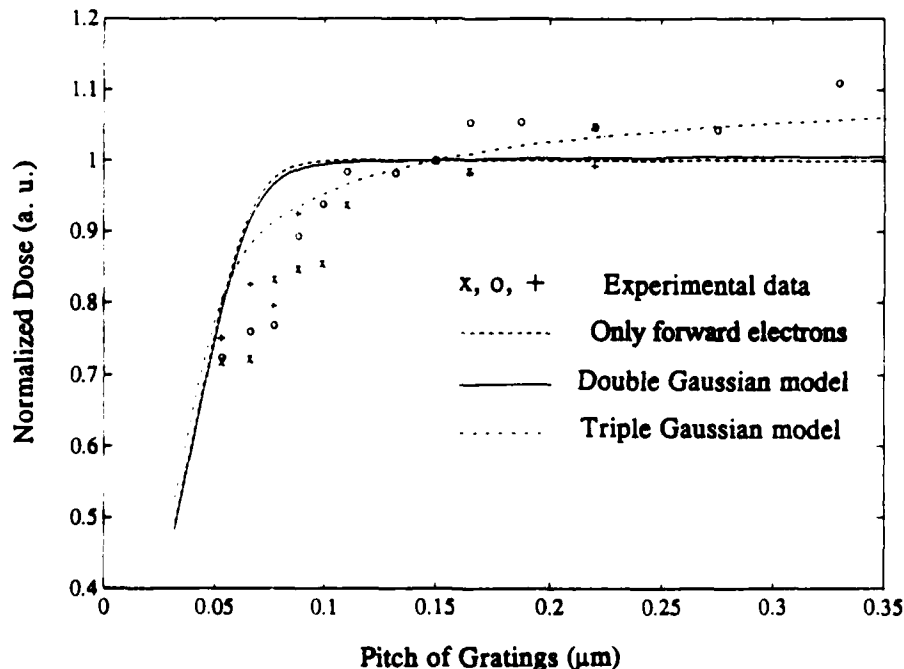


Fig. 1: Normalized dose for equal line widths as a function of grating pitch. The agreement with a triple Gaussian model shows that fast secondary electrons play a significant role in the exposure.

The strength of PMMA walls is important because the spacing of the lines is much smaller than the thickness of the starting resist. When the (positive) resist is developed out to form the lines by lift-off, a wall of PMMA remains to form the space between the two lines. Since the length of the walls is very long (on the order of 50 to 100 microns), the walls have a tendency to wave and buckle under the stress of the development process, whereas very straight walls are necessary for the experiments. This phenomenon required a very good understanding of the forces in order to control the process for the continuation of the current drag experiments. We found that the lack of stability was due not to process control or technique, but rather to the very slight swelling of the PMMA with the developer leading to buckling of the walls. Furthermore, grating structures had a tendency to fail in very regular arrays of waves which seemed at first to be due to either noise on the electron beam or poor development/drying procedures. Our study revealed that the regular patterns were in fact due to attractive forces through the developer fluid which set up a pattern of wall failure across the grating.

One new and interesting outcome of the study was the role of swelling in the development of PMMA. It was universally believed that a contributing factor to the ultra-high resolution properties of PMMA was the absence of any swelling characteristics. We found that in the regime of narrow walls, the very slight swelling that does exist as a consequence of the normal chemical development sequence of events does indeed impact the results on this size scale.

A directly applicable result of the study was the relationship that the fabrication of reliably straight walls is achieved under the condition of not exceeding a ratio of wall height to thickness of 5. This applies over a wide range of starting resist thicknesses. Using insights gained from the study, we were able to fabricate gold gratings of thickness 15 nm with pitch as small as 38 nm (Fig. 2) and gold line pairs on a pitch of 36 nm. To our knowledge, these are the densest patterns ever fabricated using direct lithographic techniques.

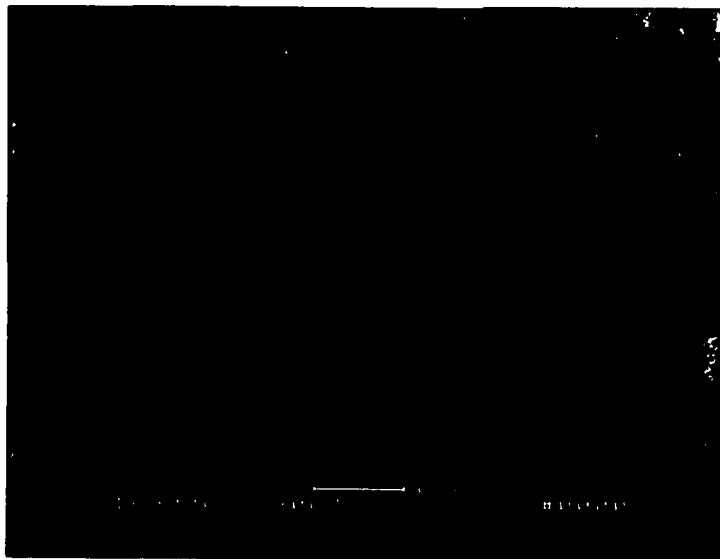


Fig. 2: Ti/Au (2 nm/15 nm) grating with 38 nm pitch fabricated by electron beam lithography using our high contrast developer.

Relevant Publications:

- G. H. Bernstein, D. A. Hill and W. P. Liu, "New High-Contrast Developers for PMMA Resist," *Journal of Applied Physics* **71**, 4066 - 4075 (1992).
- G. H. Bernstein and D. A. Hill, "On the Attainment of Optimum Development Parameters for PMMA Resist," *Superlattices and Microstructures* **11**, 237-240 (1992).
- X. Huang, G. Bazan, G. H. Bernstein, and D. A. Hill, "Stability of Thin Resist Walls," *Journal of the Electrochemical Society* **139**, 2952-2956 (1992).
- D. A. Hill, X. Huang, G. Bazan, and G. H. Bernstein, "Swelling and Surface Forces-Induced Instabilities in Nanoscopic Polymeric Structures," *Journal of Applied Physics* **72**, 4088-4094 (1992).
- X. Huang, G. H. Bernstein, G. Bazan, and D. A. Hill, "Spatial Density of Lines in PMMA by Electron Beam Lithography," *Journal of Vacuum Science and Technology A* **11**, 1739-1744 (1993).
- X. Huang, G. Bazan, G. H. Bernstein, "New Technique for Computation and Challenges for Electron Beam Lithography," *Journal of Vacuum Science and Technology B* **11**, 2665 - 2569 (1993).

Experimental Study of Coupling between Metal Lines*Investigator: G. H. Bernstein*

In order to increase confidence in our low temperature measurement techniques, we replicated previously published work by other groups in which the length dependence of universal conductance fluctuations in 1-D structures was investigated. Figure 3 shows the dependence of the magnitude of the conductance fluctuations on the length of gold wires with cross-sections of 30 nm by 70 nm at 1.8K. The existence of an L^{-n} dependence (where $n=1.5$ to 2) is characteristic of this phenomenon, and is in agreement with the previous experiments.

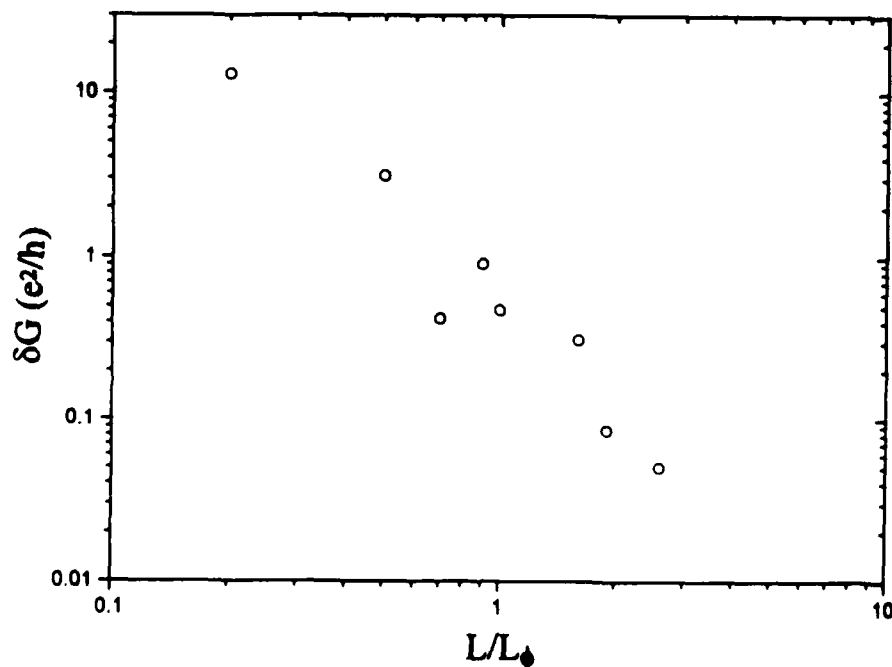


Fig. 3: Magnitude of the conductance fluctuations as a function of wire length.

Having established the credibility of our measurement set-up, we proceeded to measure current drag in normal metal wires. In this phase of the research effort, six attempts over six months were made at finding current drag between gold lines at 1.7K to 4.2K with linewidths of 50 nm and spaces of 25 nm. The decision to use linewidths larger than those discussed above was based on measurement issues. The ratio of current drag to drive current reported in the literature was on the order of 10^{-5} , and drive currents needed to be about 100 nA, putting the drag current into the pA range. In our case, we expected a considerably lower drag/drive current ratio, and did not wish to go below 10 microamps for the drive current, with the ability to measure down to 1 pA (10^{-7} ratio). Thinner lines than those used would not have withstood such large drive currents.

In the course of the experiment, considerable progress was made in improving the measurement system to 0.1 pA measurements and eliminating all sources of stray cross-talk between measurement lines, but no evidence was seen of actual coupling between the lines on chip. We believe that the reason for failure is that we had not succeeded in meeting the specifications for the patterns and measurements as outlined above. Beyond 10 nm linewidths, the screening was likely so large that any Coulomb effects directed from the drive to the drag lines were screened near the surface of the drag line, so that most of the electrons in the drag line were unaffected by the Coulombic fields induced by the drive electrons, and very little momentum transfer resulted. Given better measurement techniques now implemented, we will revisit this regime and look for possible effects.

Through extensive discussions with Dr. N. Giordano at Purdue University, we decided to pursue the work in the area of superconducting films and lines. The basis for this decision was the expectation that the Cooper pairs involved in the phenomenon of superconductivity enjoy very long coherence lengths - on the order of several tens of nm. Because of this, screening of the charge-induced momentum transfer will be much less, and therefore is expected to survive on the size scales used in the experiment.

Although he has not observed coupling between normal metal (NM) films, Giordano has already observed drag between thin films of NM and superconducting (SC) metals [private comm., unpublished]. His work showed that current drag developed within the NM and SC temperature transition regime. For this reason, he (and we) chose a material that has a wide transition temperature range so that the crossover from no drag to drag can be more easily studied. We chose aluminum/oxygen compounds which shift the critical temperature of aluminum from 1.2K to as high as 1.95K with a transition range of about 0.1K.

By introducing oxygen during aluminum evaporation, we have succeeded in achieving films with a wide range of transition temperatures. Using these films, we have observed current drag between films. Figure 4 shows induced voltage in the SC film due to current driven through a gold film. Both films are 20 nm thick, separated by 35 nm of Al_2O_3 (insulating). The coupling occurs only in the temperature range over which the SC film changes from normal to SC (solid line in the figure). It is interesting that the induced voltage decreases as the resistance of the SC film approaches zero, i.e. purely superconducting (also observed by Giordano). We believe this is due to the nature of the SC material i.e. a superconductor cannot support a voltage drop. As the temperature decreases and electrons pair up, the coherence length grows and coupling begins. However, as pairing is completed, there is a canceling of the voltage due to the coupling and the superconductivity, so the voltage decreases to zero as it must in a SC.

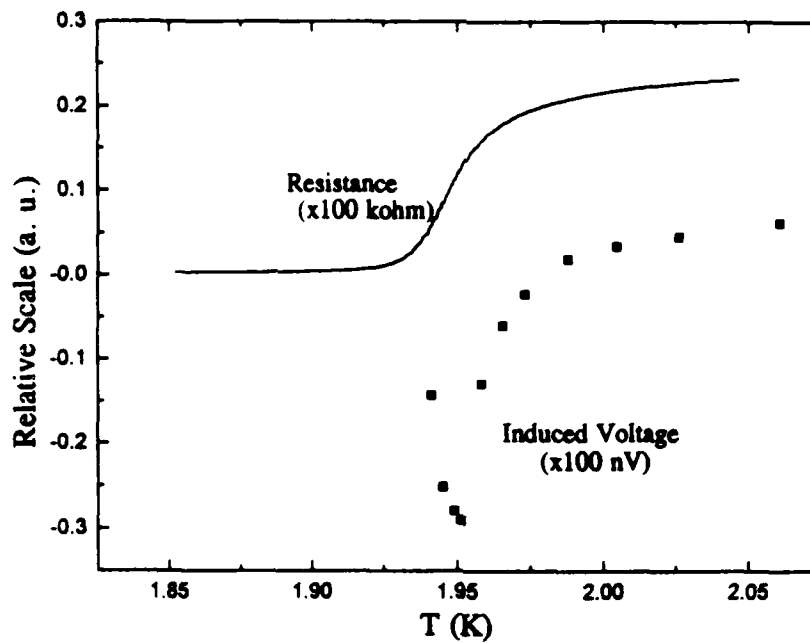


Fig. 4: Induced voltage in the superconducting film due to current driven through a gold film.

Work is continuing in order to observe coupling between NM and SC lines. We plan to pursue several schemes for current drag between lines including lines lying side-by-side and over each other. There are advantages to both schemes which is the subject of the dissertation of Xiaokang Huang. Both schemes will rely heavily on our ability to perform very accurate alignments between metal levels (see the section on Dot-Dot Coupling).

Theory of Line-Line Coupling

Investigator: S. Bandyopadhyay

A general theoretical framework was developed for studying coupling and crosstalk between ultranarrow and ultradense interconnects in ULSI. Both coupling between optical interconnects and quantum mechanical coupling between silicide interconnects were studied using this theory. It was found that generic optical interconnects are quite immune to crosstalk since GaAs/AlGaAs optical waveguides (interconnects) are non-leaky. In contrast, quantum mechanical tunneling between contiguous closely spaced silicide lines can be quite significant if the lines are about 5 nm apart. To suppress this coupling, a possible remedy is to dust the intervening dielectric with magnetic impurities which are known to inhibit tunneling. A paper describing this research was published in the *IEEE Journal of Quantum Electronics* (provided in the appendix).

Relevant Publications:

S. Bandyopadhyay, "Coupling and Crosstalk Between High Speed Interconnects in Ultra Large Scale Integrated Circuits," *IEEE Journal of Quantum Electronics* 28, 1554 - 1561 (1992).

Line - Line Coupling by Phonon-Assisted Hopping of Electrons

Investigator: S. Bandyopadhyay

A possible source of coupling between narrow contiguous wires is phonon assisted hopping of electrons from one wire to another. In addition to crosstalk, this kind of events can also cause Shot noise in the interconnects. We have calculated phonon scattering rates in quantum wires and at present are investigating phonon assisted hopping rates. We have also found that a magnetic field can dramatically suppress acoustic phonon emission in quantum wires. This also has very serious implications for the quantum Hall effect in one-dimensional systems. Several papers describing the effect of a magnetic field on phonon scattering have been published.

Relevant Publications:

- N. Telang and S. Bandyopadhyay, "Quenching of Acoustic Phonon Scattering of Electrons in Semiconductor Quantum Wires Induced by a Magnetic Field," *Applied Physics Letters* **62**, 3161 (1993).
- N. Telang and S. Bandyopadhyay, "Modulation of Electron Phonon Scattering in Quantum Wires by a Magnetic Field," *Semiconductor Science and Technology* (in press).
- N. Telang and S. Bandyopadhyay, "The Effect of a Magnetic Field on Polar Optical and Surface Phonon Scattering Rates," *Physical Review B* **48** (Dec 15 - 1993).

Magnetotransport in Quantum Wires

Investigator: S. Bandyopadhyay

A formalism was developed for calculating the electron wavefunction and the energy dispersion relation of hybrid magnetoelectric states in ultranarrow quantum wires in the presence of a magnetic field. The results are exact solutions of the Schrödinger equation. The importance of this work is that it allows one to study the effect of a magnetic field on transport in narrow quasi-one dimensional wires. We are extending this work is to study transport in two coupled wires. At present, we are studying the following problem. An electron is introduced into a wire at a certain instance of time. We would like to find out after what time the electron switches over to an adjoining wire and with what efficiency. This requires solving the time dependent Schrödinger equation, for which we have formulated a general prescription that is valid even in the presence of a magnetic field. A paper describing the first (steady-state) part of this work was published in the *Journal of Applied Physics* (provided in the appendix).

Relevant Publications:

- S. Chaudhuri and S. Bandyopadhyay, "Numerical Calculation of Hybrid Magnetoelectric States in an Electron Waveguide," *Journal of Applied Physics* **71**, 3027 - 3029 (1992).

Spatial Current Distribution in Quantum Wires

Investigator: S. Bandyopadhyay and W. Porod

A formalism was developed for studying the spatial distributions of the current, potential, excess carrier concentration, residual resistivity dipoles and electromigration forces in quantum wires in a magnetic field. A paper describing these results was published in *Superlattices and Microstructures*. It was also presented at the Sixth International Conference on Superlattices, Microstructures and Microdevices, Beijing, People's Republic of China, August, 1992. A more detailed version of this paper appeared in *Physical Review B* (provided in the appendix).

In addition, we have also obtained self-consistent electronic states through an iterative solution of the 2-D Schrödinger equation using current carrying boundary conditions. Any bound electronic charge is approximated by the semi-classical Thomas-Fermi screening model. The Hartree potential due to the electron density is explicitly determined from the total charge density in each iteration, and is used to calculate the electronic states in the subsequent iteration. The self-consistent solution is obtained when the potential difference between iterations converges to within a desired tolerance. The results of this work are described in a paper in *Superlattices and Microstructures*.

Relevant Publications:

- Henry K. Harbury, Wolfgang Porod, and Craig S. Lent, "Field Effects in Self-Consistent Transport Calculations for Narrow Split-Gate Structures," *Superlattices and Microstructures* **11**, 189 - 193 (1992).
- S. Chaudhuri and S. Bandyopadhyay, "Quantum Transport in a Disordered Quantum Wire in the Presence of a Magnetic Field," *Superlattices and Microstructures* **11**, 241 - 244 (1992).
- S. Chaudhuri and S. Bandyopadhyay, "Spatial Distribution of the Current and Fermi Carriers Around Localized Elastic Scatterers in Quantum Transport," *Physical Review B* **45**, 11126 - 11135 (1992).
- S. Bandyopadhyay, S. Chaudhuri, B. Das, and M. Cahay, "Features of quantum magnetotransport and electromigration in mesoscopic systems," *Superlattices and Microstructures* **12**, 123 - 132 (1992).
- S. Chaudhuri, S. Bandyopadhyay, and M. Cahay, "Spatial distribution of the current, Fermi carrier density, potential and electric field in a disordered quantum wire in a magnetic field," *Physical Review B* **47**, 12649 (1993).

Design of Quantum Wires at Corrugated Heterointerfaces

Investigator: W. Porod

in collaboration with Dr. S. M. Goodnick (OSU)

We have investigated the formation of a quantum wire structure by the confinement of electrons between lateral quasi-two-dimensional p-n junctions at corrugated GaAs/AlGaAs heterostructures. Such a quantum wire may be realized at the tip of a Si-doped AlGaAs overgrown V groove in a Si-GaAs substrate due to the surface orientation dependence of Si doping. The two-dimensional conduction and valence band profiles for the electron and hole charge densities are

obtained numerically within a semiclassical Thomas-Fermi screening model. The quantized electronic wire states at the heterointerface are then obtained by solving the two-dimensional effective mass Schrödinger equation using the calculated potential profile. We have explored the parameter space of the one-dimensional electronic system and we established which features of the structure are dominant factors in controlling the quantum confinement. Specifically, we found that the energy level spacing of the quantum wire depends primarily upon the lateral confinement width in the n-type region at the tip of the V groove. The ground state energy of the wire is shown to depend on both the lateral confinement width and the vertical heterointerface confinement width. We have also studied the effect of lateral gates on the side walls of the V groove in order to obtain direct control of the quantum wire transport properties. Our studies have resulted in an experimental effort in the group of Dr. Jim Merz at UC Santa Barbara, which is aimed at realizing such a quantum wire structure by MBE techniques.

Relevant Publications:

Henry K. Harbury, Wolfgang Porod, and Stephen M. Goodnick, "Lateral p-n junctions between quasi two-dimensional electron and hole systems at corrugated GaAs/AlGaAs interfaces," *Journal of Vacuum Science and Technology B* **10**, 2051 - 2055 (1992).

Wolfgang Porod, Henry K. Harbury, and Stephen M. Goodnick, "Lateral p-n junctions and quantum wires formed by quasi two-dimensional electron and hole systems at corrugated GaAs/AlGaAs interfaces," *Applied Physics Letters* **61**, 1823 - 1825 (1992).

Henry K. Harbury, Wolfgang Porod, and Stephen M. Goodnick, "A Novel Quantum Wire Formed by Lateral p-n-p Junctions Between Quasi-Two-Dimensional Electron and Hole Systems at Corrugated GaAs/AlGaAs Interfaces," *Journal of Applied Physics* **73**, 1509 - 1520 (1993).

Dot - Dot Coupling

Fabrication of dense dot arrays

Investigator: G. H. Bernstein

The simplest conception of an array of dots is to directly fabricate metal dots on the surface of a 2DEG formed by a heterostructure. This could be utilized if it is followed by connecting to a top gate after planarization by polyimide, or by reactive ion etching into the doping layer, followed by evaporation of a blanket metal layer, effecting depletion of the etched region and leaving dots in the 2DEG. We have made dot arrays as a test of our lithographic system with dot sizes approaching 10 nm on a 37 nm pitch, as shown in Fig. 5. These dots, however, are not planned as the first effort in actual dot arrays for nanostructures.

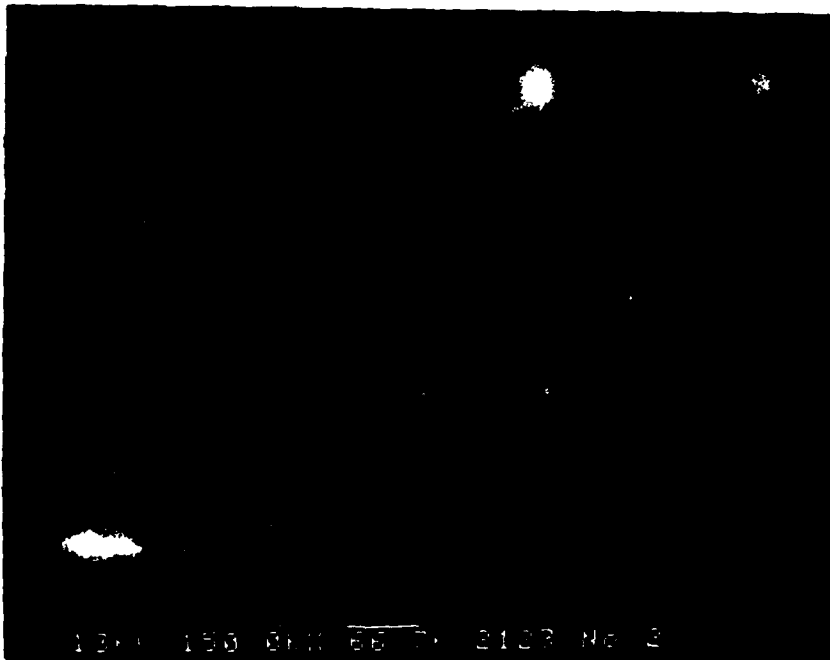


Fig. 5: Gold dots with diameters as small as 10 nm fabricated by electron beam lithography.

A more useful "dot" system is that created by single electron tunneling through small tunnel junctions onto a small metal island. We have demonstrated the processing technique of shadow evaporation with thin Al_2O_3 tunnel barriers, as described in our paper in the *Review of Scientific Instruments*. Figure 6 shows a large array of tunnel junctions using native oxide on aluminum fabricated with electron beam lithography. The overlapping areas are small, about 30 by 60 nm, but there could be significant stray capacitance due to the overlapping metal. This technique has recently been refined to decrease the stray capacitance. We have improved our layer-to-layer overlay alignment accuracy considerably. Using our new technique, described in the *Journal of Vacuum Science and Technology A*, we have achieved alignment accuracies beyond the visible limits of our newest field emission scanning electron microscope, with a resolution of about 1.5 nm.

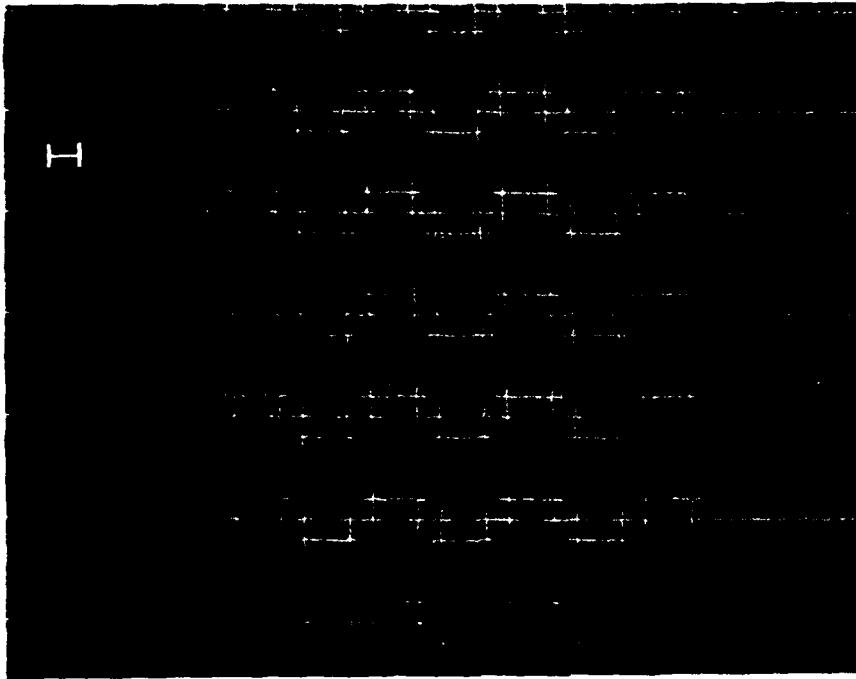


Fig. 6: Array of tunnel junctions fabricated by shadow evaporation.

Figure 7 shows a tunnel junction fabricated by the alignment of one metal line against another on subsequent processing and metalization steps. The alignment is so accurate that no horizontal offset can be seen at 250,000 times magnification. Overlap in the vertical direction exists only because of proximity effect elongation of the lines beyond their actual lithographic placement, so that with the two lines combined, an extremely small tunnel junction has resulted with an area of about 30 by 30 nm². We plan to test devices of this quality in the near future.

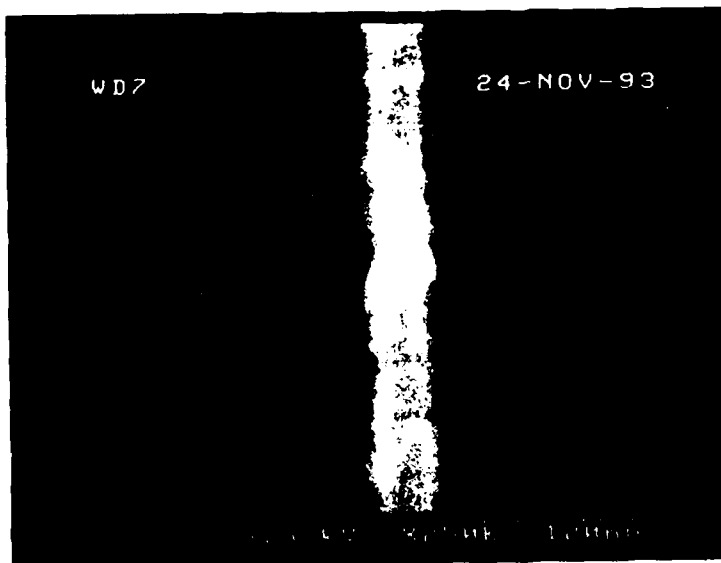


Fig. 7: Very small overlap of two wires to be used as a single electron tunnel junction. The wires were placed over each other by a nearly perfect alignment of two metal layers.

Relevant Publications:

- S. J. Koester, G. Bazan, G. H. Bernstein, and W. Porod, "Fabrication of Ultrasmall Tunnel Junctions by Electron Beam Lithography," *Review of Scientific Instruments* **63**, 1918 - 1921 (1992).
- G. Bazan and G. H. Bernstein, "Electron Beam Lithography Over Very Large Scan Fields," *Journal of Vacuum Science and Technology A* **11**, 1745-1752 (1993).

Physical Coupling between Quantum Dots for Cellular Automata Architectures

*Investigators: W. Porod and G. H. Bernstein
in collaboration with Dr. C. S. Lent*

We have studied architectures composed of quantum structures which are based on the physical coupling between the individual elements. We have pursued the general notion of computing in granular arrays of quantum dots, similar to cellular automata. In a specific example, we have implemented cellular automata rules which correspond to a solution of the Helmholtz equation. Our simulations exhibited wave behavior, such as Huygen's principle, interference, and diffraction at a slit. We have attempted to employ these wave phenomena to computing on a chip, but more work is needed.

Perhaps our most successful avenue of research was aimed at developing a specific model system for the coupling between quantum dots. In these so-called quantum cellular automata, each cell is composed of several quantum dots which are occupied by only a few electrons. Inside each cell, electrons may tunnel from one dot to the next. No tunneling is allowed between cells. Electrons, both within and between cells, interact Coulombically. We have demonstrated bistable behavior of the charge alignment in each cell, which is due to intracellular quantum mechanics and inter-cellular Coulomb coupling. This bistability is akin to gain and saturation in conventional electronics. In ongoing work, computing elements have been demonstrated, building upon the basic quantum cell.

This work has led to a new project, which is entitled "Quantum Cellular Automata," and which is sponsored by ARPA and administered by ONR.

Relevant Publications:

- Craig S. Lent, P. Douglas Tougaw, and Wolfgang Porod, "Bistable saturation in coupled quantum dots for quantum cellular automata," *Applied Physics Letters* **62**, 714 - 716 (1993).
- Craig S. Lent, P. Douglas Tougaw, Wolfgang Porod, and Gary H. Bernstein, "Quantum Cellular Automata," *Nanotechnology* **4**, 59 - 57 (1993).
- P. Douglas Tougaw, Craig S. Lent, and Wolfgang Porod, "Bistable Saturation in Coupled Quantum-Dot Cells," *Journal of Applied Physics* **74**, 3558 - 3566 (1993).

Numerical modelling of quantum dot structures

Investigator: W. Porod

Knowledge of the potential profiles in quantum dot structures is an important issue both for the

design of quantum dots as well as for a study of dot - dot coupling. We have developed a numerical technique which allows us to model the effect of close-by surface states and the interface to the dielectric. Specifically, we have studied potential distributions and quantized electronic states in gated AlGaAs/GaAs quantum dot structures obtained from self-consistent solutions of the axisymmetric Poisson and Schrödinger equations. Our model takes into account the effect of surface states by viewing the exposed surface as the interface between the semiconductor and the dielectric. We investigate the occupation of the quantum dot structures as a function of (i) gate bias, (ii) distance of the confining heterointerface from the top gate, and (iii) different physical models of the interface charge at the exposed semiconductor surface. This modelling is important for the design of our recently proposed mesoscopic computing architectures, called Quantum Cellular Automata, which consist of arrays of interacting quantum dot cells occupied by few electrons.

Relevant Publications:

Minhan Chen, Wolfgang Porod, and David J. Kirkner, "A Coupled Finite Element/Boundary Element Method for Semiconductor Quantum Devices with Exposed Surfaces," *Journal of Applied Physics* 75 (Mar 1 - 1994).

Supercomputing with spin polarized single electrons

*Investigator: S. Bandyopadhyay,
in collaboration with Drs.
Miller and Das*

We have proposed a novel quantum technology for ultrafast, ultradense and ultra-low power supercomputing. The technology utilizes *single electrons* as binary logic devices in which the *spin* of the electron encodes the bit information. The architecture mimics quantum coupled two dimensional cellular automata without physical interconnects. It is realized by laying out on a wafer regimented arrays of nanophase particles each hosting an electron. Various types of logic gates, combinational circuits for arithmetic logic units, and sequential circuits for memory can be realized. The technology has many advantages such as (1) the absence of physical interconnects between devices (inter-device interaction is provided by quantum mechanical coupling between adjacent electrons), (b) ultrafast switching times of about 1 picosecond for individual devices, (c) extremely high bit density approaching 10 Terabits/cm², (4) non-volatile memory, (5) robustness and possible room temperature operation with very high noise margin and reliability, (6) a very low power delay product for switching a single bit (about 10⁻²⁰ Joules), and (7) a very small power dissipation of a few tens of nanowatts per bit.

We are also investigating possible routes of clocking this type of a chip with coherent electromagnetic radiation that can induce magnetic dipole transitions. This is the closest realization of the Benioff-Zurek-Peres type quantum spin computer for dissipationless computation as envisaged by Feynman. We have found that the Hamiltonian of the system (including the radiation) can be Hermitian so that the computation represented by the transformation $\exp[-2\pi i H t / \hbar]$ is unitary. We are presently studying the time evolution of such a system. Finally, we have proposed a novel fabrication technology for producing single electron chips that involves selective area nucleation of nanophase particles. This work is being carried out in collaboration with Argonne National Laboratory and the National Institute of Standards and Technology, Gaithersburg, Maryland.

Line - Dot Coupling

*Investigator: W. Porod
in collaboration with Dr. C. S. Lent*

Line-to-dot coupling will be utilized in our proposed Quantum Cellular Automata architectures to sense the polarization of a particular cell. In other words, this phenomenon will serve as the READ operation. Line - dot coupling may also be used to set the initial state of an array of quantum dot cells.

So far, we have studied transmission phenomena when coupling a quantum waveguide to a resonator, which may be realized as a quantum dot. In particular, we have investigated the properties of the transmission amplitude in the complex-energy plane. We find that, similar to double-barrier resonant tunneling, there are transmission poles in the complex-energy plane for quantum waveguide structures which contain quasi-bound states in attached t-stub resonators. In contrast to double-barrier resonant tunneling, however, we also find that the quantum wire networks also possess transmission zeros (antiresonances), which always occur on the real-energy axis. The existence of transmission zeros is a characteristic feature of quantum waveguide system with attached resonators, but is absent for double-barrier resonant tunneling, which contains the resonant cavity as part of the transmission channel. We demonstrate that each quasi-bound state of the resonantly-coupled quantum waveguide system leads to a zero-pole pair of the transmission amplitude in the complex-energy plane. The previously noted resonance - antiresonance behavior of the transmission probability, which leads to its sharp variation as a function of energy, can be understood in terms of these zero-pole pairs.

We have also investigated the line shape of the transmission probability in quantum waveguides with resonantly-coupled cavities. Resonance/antiresonance features in the vicinity of each quasi-bound state can be characterized by a zero-pole pair in the complex-energy plane, which leads to asymmetrical transmission peaks. We have found a generalization of the familiar symmetrical Lorentzian line. Using several examples, we demonstrate the utility of our proposed line shape to extract the lifetime of the quasi-bound state by a fit to the data. We also discuss the asymmetrical line shapes in the context of Fano resonances.

Relevant Publications:

Wolfgang Porod, Zhi-an Shao, and Craig S. Lent, "Transmission Resonances and Zeros in Quantum Waveguides with Resonantly-Coupled Cavities," *Applied Physics Letters* **61**, 1350 - 1352 (1992).

Wolfgang Porod, Zhi-an Shao, and Craig S. Lent, "Resonance-Antiresonance Line Shape for Transmission in Quantum Waveguides with Resonantly-Coupled Cavities," *Physical Review B* **48**, 8495 - 8498 (1993).

Zhi-an Shao, Wolfgang Porod, and Craig S. Lent, "Transmission Resonances and Zeros in Quantum Waveguide Systems with Attached Resonators," *Physical Review B* (in press).

Photovoltaic Effect in the Presence of Magnetic Fields

*Investigator: G. H. Bernstein
in collaboration with Dr. N. Giordano (Purdue)*

From a practical point of view, metal interconnects used in mesoscopic systems will be subject to some electrical noise and interference. Since circuits utilizing mesoscopic phenomena will operate at frequencies high in the microwave range, we must be concerned with the effects of microwave bombardment and current drive on mesoscopic systems. Of a scientific interest, the existence of photovoltages induced by bombardment of microwave radiation (the photovoltaic (PV) effect) has proved fertile ground for studying the interaction of high frequency electric fields on conduction electrons in metals. One present model for the PV effect is that electrons absorb microwave photons, and being out of equilibrium with the Fermi sea, the energetic electrons diffuse in a direction preferential to the precise impurity distribution. Because of this, a DC voltage on the order of nV develops across the leads of the device.

We have engaged in an extensive collaboration with Dr. Giordano and graduate student R. Bartolo to study the photovoltaic effect in mesoscopic structures. Prior to our involvement, their work was confined to structures with dimensions of about one micron. We have collaborated to produce a variety of structures with dimensions to 50 nm in the form of wires and rings of diameter 330 to 500 nm, measured at Purdue University.

We have fabricated (on glass substrates) gold 1-D wires with widths of 50 to 70 nm, thicknesses of 20 nm, and total lengths of up to 1.5 microns. These structures clearly showed evidence of the PV effect at 4.2K. Although the results from the small wires were essentially the same as the larger structures, the small structures allowed comparisons with similarly sized lines formed in the shape of rings for investigation of the Aharonov-Bohm (A-B) effect. In these structures, A-B oscillations in the PV data were clearly visible. Figure 8a shows the data for a 330 nm diameter

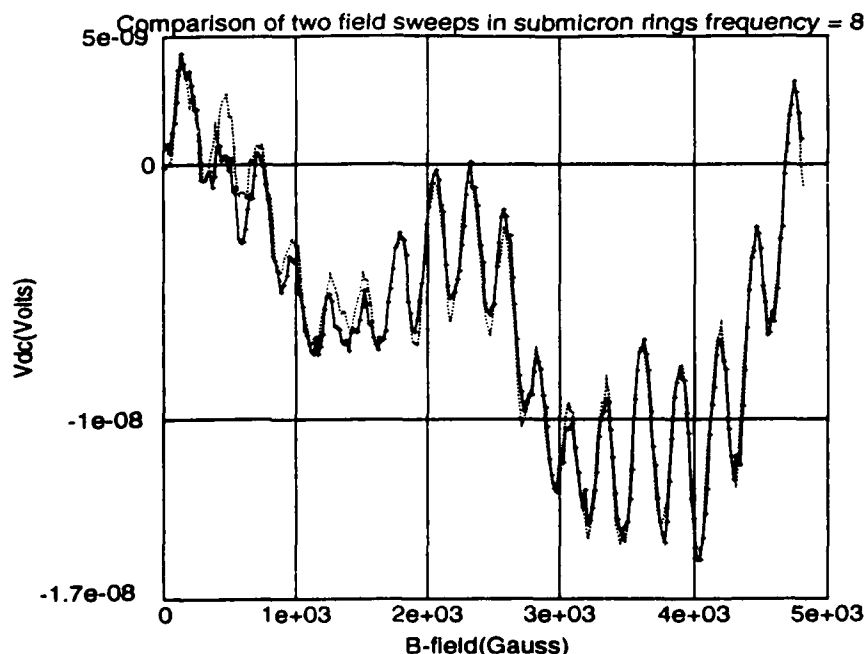


Fig. 8a: Photovoltaic effect in a gold ring of 330 nm diameter. The two lines are opposite sweeps of the magnetic field.

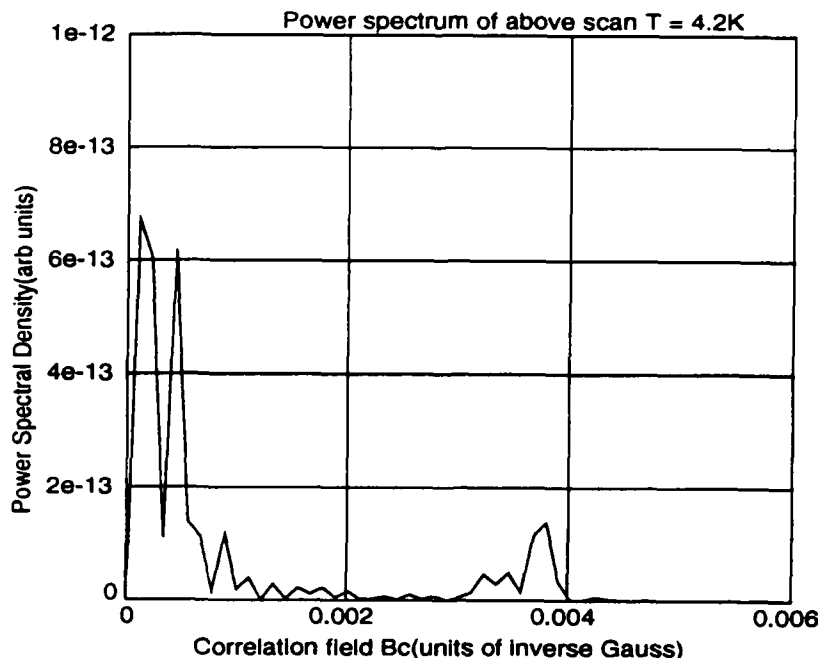


Fig. 8b: Fourier transform of the data in Fig. 8a. The peak corresponds to h/e oscillations in the photovoltaic effect indicating a phase coherent component to the phenomenon.

ring at 4.2K and 8.4 GHz radiation under a sweep of the magnetic field up to 1 Tesla. The data has been shown to be very consistent with time and sweep direction. Fig. 8b shows the Fourier transform in magnetic field with a clear peak at about 0.0037/Tesla. This peak is at precisely the correct location corresponding to h/e oscillations. The data indicate that the existence of the PV effect is in fact mesoscopic in nature, and is related to the phase coherence over a large fraction of one micron. This is the first data that unambiguously shows this fact.

Another interesting point is that previous observations of the A-B effect have been performed at temperatures of less than 0.1K. Here we have demonstrated the direct observation of an A-B related phenomenon at much higher temperatures. This could prove useful in future investigations of the A-B effect and other physical phenomena.

This work has been presented at the 1992 40th Midwest Solid State Conference and will be presented at the 1994 March Meeting of the APS. A publication is under preparation.

A New Technique for Fabricating Mesoscopic Structures in Silicon

Investigators: G. H. Bernstein and W. Porod

In the interest of fabricating wires with new properties and in new materials, we have investigated a novel technique for fabricating mesoscopic structures in general, and quantum wires in particular, in silicon. Initially, we performed a series of experiments to investigate the spatial stability of positive charge induced in SiO_2 by exposure to a beam of electrons. The positive charge caused a shift in the threshold voltage of MOSFETs by up to -15 volts, so that the area exposed to the beam inverted, creating a 2DEG in the silicon, at a much more negative voltage than did the area left unexposed.

We showed through the use of capacitance-voltage measurements, as discussed in our paper in *Scanning*, that the induced positive charge remains stable over periods of at least several weeks. Furthermore, the charge did not appear to move in a measurable way during the testing period. We surmised that by intentionally introducing positive charges into the oxide of a MOSFET in a controlled manner using electron beam lithography techniques, that quantum structures could be fabricated.

We have so far investigated various aspects of producing 1D structures with this technique. We made MOSFETs in our laboratory with gate lengths and widths of 8 microns and 10 microns, respectively. After careful evaluation, the devices were exposed by the electron beam in a single line from source to drain. At room temperature, the device I-V curves showed a slight enhancement of the drain current at voltages slightly lower than the original threshold voltage. We infer that this was due to the narrow line of positive charge.

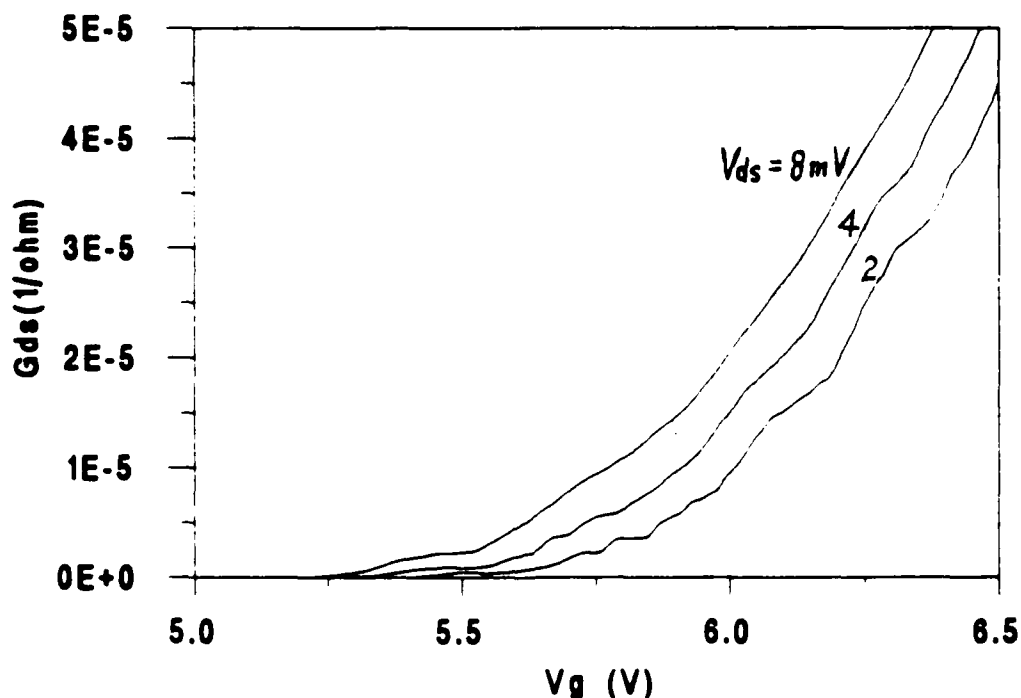


Fig. 9: Conductance in a MOSFET exposed from source to drain by a narrow electron beam. The steps in the conductance are due to the formation of a quantized channel.

We tested several devices, both exposed and unexposed at 1.8K. Among the unexposed devices, the Notre Dame devices and commercially fabricated devices showed no structure whatsoever around the threshold voltages. All of the exposed devices showed complicated structure very close to threshold. One set of conductance data as a function of the gate voltage for different drain/source voltages is shown in Fig. 9. The steps indicate that right at threshold, the line does not invert uniformly, leaving a pinched area or areas in series with the resistance of the channel. Steps are clearly evident, although they do not appear in units of $e^2/4h$, as they ought to for point contacts in silicon. We concluded that the lack of predictable step size in the conductance was due to the large series resistance of the 1D channel. We are in the process of performing more experiments on this system and are preparing a manuscript regarding the preliminary results.

Relevant Publications:

G. H. Bernstein, S. W. Polchlopek, R. Kamath, and W. Porod, "Determination of Fixed Electron-Beam-Induced Positive Oxide Charge," *Scanning* **14**, 345 - 349 (1992).

Journal Publications

- * Items marked with an asterisk are included in the appendix.
- * S. Bandyopadhyay, "Coupling and Crosstalk Between High Speed Interconnects in Ultra Large Scale Integrated Circuits," *IEEE Journal of Quantum Electronics* **28**, 1554 - 1561 (1992).
- * G. H. Bernstein, D. A. Hill and W. P. Liu, "New High-Contrast Developers for PMMA Resist," *Journal of Applied Physics* **71**, 4066 - 4075 (1992).
- * S. J. Koester, G. Bazan, G. H. Bernstein, and W. Porod, "Fabrication of Ultrasmall Tunnel Junctions by Electron Beam Lithography," *Review of Scientific Instruments* **63**, 1918 - 1921 (1992).
- S. Subramaniam, S. Bandyopadhyay, B. Das, G. H. Bernstein, and P. A. Sekula-Moise, "Anomalous Magnetoresistance Due to Correlations Between Disorder and Boundary Scatterings in a Two-Dimensional Electron Gas Channel," *Physical Review B* **45**, 3826 - 3829 (1992).
- N. Telang and S. Bandyopadhyay, "A Monte Carlo Study of Correlations Between Impurity Scattering Events in a Two dimensional Electron Gas Causing Inhomogeneous Magnetoresistance," *Superlattices and Microstructures* **11**, 99 - 102 (1992).
- * Henry K. Harbury, Wolfgang Porod, and Craig S. Lent, "Field Effects in Self-Consistent Transport Calculations for Narrow Split-Gate Structures," *Superlattices and Microstructures* **11**, 189 - 193 (1992).
- * G. H. Bernstein and D. A. Hill, "On the Attainment of Optimum Development Parameters for PMMA Resist," *Superlattices and Microstructures* **11**, 237-240 (1992).
- S. Chaudhuri and S. Bandyopadhyay, "Quantum Transport in a Disordered Quantum Wire in the Presence of a Magnetic Field," *Superlattices and Microstructures* **11**, 241 - 244 (1992).
- S. Chaudhuri and S. Bandyopadhyay, "Spatial Distribution of the Current and Fermi Carriers Around Localized Elastic Scatterers in Quantum Transport," *Physical Review B* **45**, 11126 - 11135 (1992).
- * S. Chaudhuri and S. Bandyopadhyay, "Numerical Calculation of Hybrid Magnetoelectric States in an Electron Waveguide," *Journal of Applied Physics* **71**, 3027 - 3029 (1992).
- Henry K. Harbury, Wolfgang Porod, and Stephen M. Goodnick, "Lateral p-n junctions between quasi two-dimensional electron and hole systems at corrugated GaAs/AlGaAs interfaces," *Journal of Vacuum Science and Technology B* **10**, 2051 - 2055 (1992).
- * Wolfgang Porod, Zhi-an Shao, and Craig S. Lent, "Transmission Resonances and Zeros in Quantum Waveguides with Resonantly-Coupled Cavities," *Applied Physics Letters* **61**, 1350 - 1352 (1992).
- * Wolfgang Porod, Henry K. Harbury, and Stephen M. Goodnick, "Lateral p-n junctions and quantum wires formed by quasi two-dimensional electron and hole systems at corrugated GaAs/AlGaAs interfaces," *Applied Physics Letters* **61**, 1823 - 1825 (1992).
- * X. Huang, G. Bazan, G. H. Bernstein, and D. A. Hill, "Stability of Thin Resist Walls," *Journal of the Electrochemical Society* **139**, 2952-2956 (1992).
- * G. H. Bernstein, S. W. Polchlopek, R. Kamath, and W. Porod, "Determination of Fixed Electron-Beam-Induced Positive Oxide Charge," *Scanning* **14**, 345 - 349 (1992).

-
- * D. A. Hill, X. Huang, G. Bazan, and G. H. Bernstein, "Swelling and Surface Forces-Induced Instabilities in Nanoscopic Polymeric Structures," *Journal of Applied Physics* **72**, 4088-4094 (1992).
 - * S. Bandyopadhyay, S. Chaudhuri, B. Das, and M. Cahay, "Features of quantum magnetotransport and electromigration in mesoscopic systems," *Superlattices and Microstructures* **12**, 123 - 132 (1992).
 - L. Rota, F. Rossi, S. M. Goodnick, P. Lugli, E. Molinari, and W. Porod, "Reduced carrier cooling and thermalization in semiconductor quantum wires," *Physical Review B* **47**, 1632 - 1635 (1993).
 - N. Telang and S. Bandyopadhyay, "The effect of collision retardation on hot electron transport in a quantum well," *Physical Review B* **47**, 9900 (1993).
 - S. W. Polchlopek, G. H. Bernstein, and R. Kwor, "Properties of Nitrogen Implanted SOI Substrates," *IEEE Transactions on Electron Devices* **40**, 385 - 391 (1993).
 - * S. Chaudhuri, S. Bandyopadhyay, and M. Cahay, "Spatial distribution of the current, Fermi carrier density, potential and electric field in a disordered quantum wire in a magnetic field," *Physical Review B* **47**, 12649 (1993).
 - * Craig S. Lent, P. Douglas Tougaw, and Wolfgang Porod, "Bistable saturation in coupled quantum dots for quantum cellular automata," *Applied Physics Letters* **62**, 714 - 716 (1993).
 - * Craig S. Lent, P. Douglas Tougaw, Wolfgang Porod, and Gary H. Bernstein, "Quantum Cellular Automata," *Nanotechnology* **4**, 59 - 57 (1993).
 - N. Telang and S. Bandyopadhyay, "Quenching of Acoustic Phonon Scattering of Electrons in Semiconductor Quantum Wires Induced by a Magnetic Field," *Applied Physics Letters* **62**, 3161 (1993).
 - * Henry K. Harbury, Wolfgang Porod, and Stephen M. Goodnick, "A Novel Quantum Wire Formed by Lateral p-n-p Junctions Between Quasi-Two-Dimensional Electron and Hole Systems at Corrugated GaAs/AlGaAs Interfaces," *Journal of Applied Physics* **73**, 1509 - 1520 (1993).
 - * P. Douglas Tougaw, Craig S. Lent, and Wolfgang Porod, "Bistable Saturation in Coupled Quantum-Dot Cells," *Journal of Applied Physics* **74**, 3558 - 3566 (1993).
 - * Wolfgang Porod, Zhi-an Shao, and Craig S. Lent, "Resonance-Antiresonance Line Shape for Transmission in Quantum Waveguides with Resonantly-Coupled Cavities," *Physical Review B* **48**, 8495 - 8498 (1993).
 - * X. Huang, G. H. Bernstein, G. Bazan, and D. A. Hill, "Spatial Density of Lines in PMMA by Electron Beam Lithography," *Journal of Vacuum Science and Technology A* **11**, 1739-1744 (1993).
 - * G. Bazan and G. H. Bernstein, "Electron Beam Lithography Over Very Large Scan Fields," *Journal of Vacuum Science and Technology A* **11**, 1745-1752 (1993).
 - * X. Huang, G. Bazan, G. H. Bernstein, "New Technique for Computation and Challenges for Electron Beam Lithography," *Journal of Vacuum Science and Technology B* **11**, 2665 - 2569 (1993).
 - N. Telang and S. Bandyopadhyay, "The Effect of a Magnetic Field on Polar Optical and Surface Phonon Scattering Rates," *Physical Review B* **48** (Dec 15 - 1993).
-

- Minhan Chen, Wolfgang Porod, and David J. Kirkner, "A Coupled Finite Element/Boundary Element Method for Semiconductor Quantum Devices with Exposed Surfaces," *Journal of Applied Physics* 75 (Mar 1 - 1994).
- M. Cahay and S. Bandyopadhyay, "Quantum Devices" *Advances in Electron Physics* (Invited, in press).
- N. Telang and S. Bandyopadhyay, "Modulation of Electron Phonon Scattering in Quantum Wires by a Magnetic Field," *Semiconductor Science and Technology* (in press).
- Zhi-an Shao, Wolfgang Porod, and Craig S. Lent, "Transmission Resonances and Zeros in Quantum Waveguide Systems with Attached Resonators," *Physical Review B* (in press).
- G. H. Bernstein, R. J. Minniti, Jr., and X. Huang, "An Advanced IC Processing Laboratory at the University of Notre Dame," *IEEE Transactions on Education* (in press).

Conference Publications and Presentations

- N. M. Jisrawi, S. J. Koester, Q. Lu, M. J. Honkanen, W. Porod, G. H. Bernstein, and S. T. Ruggiero, "Microwave Studies of Single Electron Charging Effects," abstract published in the *Bulletin of the APS*, Vol. 36, No. 3, p. 872, March 1991; presented at the *1991 March Meeting of the American Physical Society*, Cincinnati, Ohio, March 1991.
- H. K. Harbury, W. Porod, and C. S. Lent, "Self-Consistent Calculation of Coherent Transport in Split-Gate Structures, abstract published in the *Bulletin of the APS*, Vol. 36, No. 3, p. 1049, March 1991; presented at the *1991 March Meeting of the American Physical Society*, Cincinnati, Ohio, March 1991.
- W. Porod, (Invited), "A Monte Carlo Model of High - Field Electronic Transport and Hot Electron Trapping/Detrapping in SiO_2 ," *Proceedings of the Seventh International Conference on the Numerical Analysis of Semiconductor Devices and Integrated Circuits*, NASECODE-7, Ed. by JJH Miller, pp. 81 - 82; presented at the *Seventh International Conference on the Numerical Analysis of Semiconductor Devices and Integrated Circuits*, NASECODE-7, Copper Mountain, Colorado, April 1991.
- H. K. Harbury, W. Porod, and C. S. Lent, "Self-Consistent Calculation of Coherent Transport in Split-Gate Structures," *Proceedings of the Seventh International Conference on the Numerical Analysis of Semiconductor Devices and Integrated Circuits*, NASECODE-7, Ed. by JJH Miller, pp. 136 - 137; presented at the *Seventh International Conference on the Numerical Analysis of Semiconductor Devices and Integrated Circuits*, NASECODE-7, Copper Mountain, Colorado, April 1991.
- Wolfgang Porod, (Invited), "Computation in Granular Flatland: Discrete Space -- Time Cellular Automata," presented at the *Granular Nanoelectronics Workshop*, Urbana, Illinois, May 1991.
- Henry K. Harbury, Wolfgang Porod, and Craig S. Lent, "Local-Field Effects in Self-Consistent Transport Calculations for Narrow Split-Gate Structures," presented at the *International Symposium on Nanostructures and Mesoscopic Systems*, Santa Fe, New Mexico, May 1991.
- G. H. Bernstein and D. A. Hill, "On the Attainment of Optimum Development Parameters for PMMA Resist," poster presented at the *International Symposium on Nanostructures and Mesoscopic Systems*, Santa Fe, New Mexico, May 1991.
- Wolfgang Porod, (Invited), "Locally - Interconnected Architectures," presented at the *DSRC/DARPA Workshop on Cellular Automata and Quantum Devices*, La Jolla, California, July 1991.
- Henry Harbury, Wolfgang Porod, and Stephen M. Goodnick, "Lateral p-n junctions between quasi two-dimensional electron and hole systems at corrugated GaAs/AlGaAs interfaces," presented at the *19th International Conference on the Physics and Chemistry of Semiconductor Interfaces*, PCSI-19, Death Valley, January 1992.
- Paul D. Tougaw, Craig S. Lent, Wolfgang Porod and G. H. Bernstein, "Cellular automata rules derived from quantum dot coupling," abstract published in the *Bulletin of the APS*, Vol. 37, No. 1, p. 135, March 1992; presented at the *1992 March Meeting of the American Physical Society*, Indianapolis, Indiana, March 1992.
- Henry K. Harbury, Wolfgang Porod, and Stephen M. Goodnick, "Lateral p-n junctions between quasi two dimensional electron and hole systems at corrugated GaAs/AlGaAs interfaces,"

- abstract published in the *Bulletin of the APS*, Vol. 37, No. 1, p. 324, March 1992; presented at the 1992 March Meeting of the American Physical Society, Indianapolis, Indiana, March 1992.
- Wolfgang Porod, Henry K. Harbury, and Gary H. Bernstein, "A Study of Cellular Automata Architectures for Integrated Quantum Device Arrays Using the Helmholtz Equations as a Computational Model," abstract published in the *Bulletin of the APS*, Vol. 37, No. 1, pp. 134 -- 135, March 1992; presented at the 1992 March Meeting of the American Physical Society, Indianapolis, Indiana, March 1992.
- Zhi-an Shao, Wolfgang Porod, and Craig S. Lent, "Transmission Resonances and Zeros in Quantum Waveguide Structures," abstract published in the *Bulletin of the APS*, Vol. 37, No. 1, pp. 289 -- 290, March 1992; presented at the 1992 March Meeting of the American Physical Society, Indianapolis, Indiana, March 1992.
- Lucio Rota, Stephen M. Goodnick, Wolfgang Porod, Fausto Rossi, and Paolo Lugli, "Electron-Electron Interaction in Multisubband Quantum Wires," abstract published in the *Bulletin of the APS*, Vol. 37, No. 1, p. 239, March 1992; presented at the 1992 March Meeting of the American Physical Society, Indianapolis, Indiana, March 1992.
- S. Chaudhuri, S. Bandyopadhyay and M. Cahay, "Spatial distribution of current and the Fermi density of states around impurities in phase coherent transport," presented at the 1992 March Meeting of the American Physical Society, Indianapolis, Indiana, March 1992.
- S. Bandyopadhyay, S. Chaudhuri and M. Cahay, "Two dimensional current and probability density profiles in mesoscopic systems in the presence of a magnetic field," presented at the 1992 March Meeting of the American Physical Society, Indianapolis, Indiana, March 1992.
- S. Chaudhuri, S. Bandyopadhyay and M. Cahay, "Study of universal conductance fluctuations and quenching of Anderson localization in a magnetic field: a Weber function analysis," presented at the 1992 March Meeting of the American Physical Society, Indianapolis, Indiana, March 1992.
- S. Bandyopadhyay, "Electromagnetic and quantum mechanical coupling between narrow interconnects in ULSI," presented at the 1992 March Meeting of the American Physical Society, Indianapolis, Indiana, March 1992.
- N. Telang and S. Bandyopadhyay, "A Monte Carlo study of correlations between impurity scattering events giving rise to anomalous magnetoresistance in a 2 DEG," presented at the 1992 March Meeting of the American Physical Society, Indianapolis, Indiana, March 1992.
- N. Telang and S. Bandyopadhyay, "The effect of finite collision duration on hot electron transport in a quantum well," presented at the 1992 March Meeting of the American Physical Society, Indianapolis, Indiana, March 1992.
- S. Subramaniam, S. Bandyopadhyay, B. Das, and G. H. Bernstein, "Anomalous magnetoresistance due to correlations between scattering events in an InGaAs/AlGaAs quantum well," presented at the 1992 March Meeting of the American Physical Society, Indianapolis, Indiana, March 1992.
- X. Huang, S. Subramaniam, G. H. Bernstein, and S. Bandyopadhyay, "Nanolithographic Fabrication of Narrow Metal Lines," presented at the 1992 March Meeting of the American Physical Society, Indianapolis, Indiana, March 1992.
- G. H. Bernstein and X. Huang, "Fabrication of Interconnect Structures for Quantum Based Archi-

- tections," presented at the *1992 March Meeting of the American Physical Society*, Indianapolis, Indiana, March 1992.
- B. Das, S. Subramaniam, S. Bandyopadhyay, and G. H. Bernstein, "Electron beam damage during electron beam lithography in modulation doped heterostructures," presented at the *1992 March Meeting of the American Physical Society*, Indianapolis, Indiana, March 1992.
- L. Rota, F. Rossi, P. Lugli, E. Molinari, S. M. Goodnick, and W. Porod, "Monte Carlo Simulation of a 'True' Quantum Wire," presented at the *1992 Symposium on Compound Semiconductor Physics and Devices*, of SPIE - The International Society of Optical Engineering, Somerset, New Jersey, March 1992.
- M. Cahay, T. Singh and S. Bandyopadhyay, "Electron emission from a quantum well as a result of exchange and Coulomb interactions," *Proceedings of the International Workshop on Computational Electronics*, pp. 147 - 150; presented at the *International Workshop on Computational Electronics*, Urbana, Illinois, May 1992.
- Craig S. Lent, Douglas Tougaw, and Wolfgang Porod, "A Bistable Quantum Cell for Cellular Automata," *Proceedings of the International Workshop on Computational Electronics*, pp. 163 - 166; presented at the *International Workshop on Computational Electronics*, Urbana, Illinois, May 1992.
- N. Telang and S. Bandyopadhyay, "Electron phonon scattering in quantum wires subjected to high magnetic field," *Proceedings of the International Workshop on Computational Electronics*, 237 - 240; presented at the *International Workshop on Computational Electronics*, Urbana, Illinois, May 1992.
- Henry K. Harbury, Wolfgang Porod, and Stephen M. Goodnick, "A Numerical Study of Lateral p-n Junctions between Quasi Two-Dimensional Electron and Hole Systems at Corrugated GaAs/AlGaAs Interfaces," *Proceedings of the International Workshop on Computational Electronics*, pp. 249 - 251; presented at the *International Workshop on Computational Electronics*, Urbana, Illinois, May 1992.
- Zhi-an Shao, Wolfgang Porod, and Craig S. Lent, "A Numerical Study of Transmission Resonances and Zeros in Quantum Waveguide Structures," *Proceedings of the International Workshop on Computational Electronics*, pp. 253 - 256; presented at the *International Workshop on Computational Electronics*, Urbana, Illinois, May 1992.
- S. Chaudhuri, S. Bandyopadhyay and M. Cahay, "Numerical study of quantum magnetotransport in disordered non-adiabatic constrictions," *Proceedings of the International Workshop on Computational Electronics*, 305 - 308; presented at the *International Workshop on Computational Electronics*, Urbana, Illinois, May 1992.
- S. Bandyopadhyay, S. Chaudhuri, B. Das, and M. Cahay, (Invited), "Magnetotransport and electromigration in mesoscopic systems," *Sixth International Conference on Superlattices and Microstructures*, Xian, People's Republic of China, August 1992.
- R. E. Bartolo, N. Giordano, X. Huang, and G. H. Bernstein, "Giant Oscillations in Mesoscopic Photovoltaic Effect," poster presented at the *40th Midwest Solid State Conference* (Judged Second-Best Poster at Conference.), Urbana, Illinois, October 1992.
- G. H. Bernstein, D. A. Hill, X. Huang, and G. Bazan, "Failure Mechanisms of Very Narrow

- PMMA Walls," presented at the *39th National Symposium of the American Vacuum Society*, Chicago, Illinois, November 1992.
- G. Bazan and G. H. Bernstein, "Electron Beam Lithography Over Large Scan Fields," presented at the *39th National Symposium of the American Vacuum Society*, Chicago, Illinois, November 1992.
- R. Frankovic and G. H. Bernstein, "Observation of Current-Induced Failures of Ultra-Narrow Metal Interconnects," poster presented at the *39th National Symposium of the American Vacuum Society*, Chicago, Illinois, November 1992.
- Craig S. Lent, P. Douglas Tougaw, Wolfgang Porod, and Gary Bernstein, "Quantum Cellular Automata," presented at the *International Symposium on New Phenomena in Mesoscopic Structures*, Maui, Hawaii, December 1992.
- Henry K. Harbury, Wolfgang Porod, and Stephen M. Goodnick, "Novel Quantum Wire Formed by Lateral p-n-p Junctions Between Quasi-Two-Dimensional Electron and Hole Systems at Corrugated GaAs/AlGaAs Interfaces," presented at the *International Workshop on Quantum Structures*, Santa Barbara, California, March 1993.
- Craig S. Lent, P. Douglas Tougaw, Wolfgang Porod, and Gary Bernstein, "Quantum Cellular Automata," presented at the *International Workshop on Quantum Structures*, Santa Barbara, California, March 1993.
- S. Bandyopadhyay, B. Das, M. Cahay and S. Chaudhuri, "Low Temperature Conduction in Ultranarrow Wires: Quantum Transport and Weak Electromigration Causing $1/f$ Noise," presented at the *183rd Meeting of the Electrochemical Society*, Honolulu, Hawaii, May 1993.
- N. Telang and S. Bandyopadhyay, "Influence of Phonons on Electromigration in Ultranarrow Wires," presented at the *183rd Meeting of the Electrochemical Society*, Honolulu, Hawaii, May 1993.
- Gary H. Bernstein, Xiaokan Huang, Greg Bazan, Davide A. Hill, Craig S. Lent, and Wolfgang Porod, "New Technique for Computation and Challenges for Electron Beam Lithography," presented at the *Three-Beam Conference*, San Diego, California, June 1993.
- Henry K. Harbury and Wolfgang Porod, "Numerical Linear Response Study of the Electron Wind Force of Electromigration in Quasi-1D Transport," presented at the *International Workshop on Computational Electronics*, Leeds, England, August 1993.
- Minhan Chen, Wolfgang Porod, and David J. Kirkner, "Boundary Conditions for Quantum Devices with Exposed Surfaces," presented at the *International Workshop on Computational Electronics*, Leeds, England, August 1993.
- Zhi-an Shao, Wolfgang Porod, and Craig S. Lent, "Resonance/Antiresonance Lineshape for Transmission in Quantum Waveguides with Resonantly-Coupled Cavities," presented at the *International Workshop on Computational Electronics*, Leeds, England, August 1993.
- Craig S. Lent, Wolfgang Porod, and P. Douglas Tougaw, "Quantum Simulation of Several-Particle Systems," presented at the *International Workshop on Computational Electronics*, Leeds, England, August 1993.
- N. Telang and S. Bandyopadhyay, "Modulation of Electron-Phonon Scattering Rates in Quantum

Wires by a Magnetic Field," presented at the *International Conference on Hot Carriers in Semiconductors*, Oxford, England, August 1993.

Craig S. Lent, P. Douglas Tougaw, Wolfgang Porod, and Gary H. Bernstein, (Invited), "Quantum Cellular Automata," presented at the *21st Midwest Solid State Theory Symposium*, Detroit, Michigan, October 1993.

Craig S. Lent, P. Douglas Tougaw, Wolfgang Porod, and Gary H. Bernstein, (Invited), "Quantum Cellular Automata," presented at the *ARPA ULTRA Program Review*, Santa Fe, New Mexico, October 1993.

Coupling and Crosstalk Between High Speed Interconnects in Ultralarge Scale Integrated Circuits

Supriyo Bandyopadhyay, *Senior Member, IEEE*

Abstract—The advent of sophisticated lithographic techniques has made it possible to fabricate densely packed ultralarge-scale-integrated (ULSI) circuits. In these chips, interconnect lines are so narrow and spaced in such close proximity that signal from one line could easily get coupled to another causing interference and crosstalk. This paper presents a general theory to model coupling between optical interconnects (waveguides) and quantum-mechanical coupling between narrow and very closely spaced silicide interconnects embedded in dielectrics (SiO_2).

I. INTRODUCTION

RECENT advances in nanolithography [1]–[6] have made it possible to delineate electronic devices with feature sizes of a few hundred angstroms in both silicon and GaAs wafers. The ability to make such small features has led to the development of ultralarge-scale-integrated (ULSI) circuits with packing densities far greater than those of very-large-scale-integrated (VLSI) circuits. In 1980, Hewlett-Packard produced a single chip microprocessor with 0.5 million devices on a 1 sq cm chip using 1.25 μm feature sizes [7]. Today, we are envisioning chips with 10^9 devices. Simple scaling law shows that the feature sizes required for such ULSI chips will be a few hundred angstroms causing extremely dense packing.

In ULSI chips not only are the devices densely packed, but so are the interconnects. The close proximity of interconnect lines resulting from this dense packing could lead to severe problems with mutual interference and crosstalk. In this paper, we study crosstalk due to coupling between optical interconnects (GaAs waveguides embedded in AlGaAs) and coupling between narrow and closely spaced silicide interconnects surrounded by a dielectric. In the case of silicide interconnects, coupling is caused by quantum-mechanical tunneling of electrons from one line to another. This kind of coupling will be especially severe if the dielectric in which the interconnects are embedded is leaky (such as porous Si_3N_4 or SiO_2

grown by wet oxidation). In addition to causing crosstalk, tunneling can give rise to a unique problem. In multilayered interconnects, there can be crossings of two lines with a very thin dielectric layer sandwiched between them. If the thickness of the lines is a few hundred angstroms, then at the crossing we have a crossover capacitor whose effective plate area is a few hundred angstroms square and the plate separation is also of the same order. The corresponding capacitance can be estimated from standard formula [8]. If the linewidths are 300 Å and the plate separation is 100 Å, then the crossover capacitance is $\sim 10^{-17}$ F. Since the dielectric layer between the plates is thin enough, an electron can tunnel through this layer from one interconnect to another. Such tunneling can charge up the capacitor to 10 mV per electron! Effects such as these have already been observed in single discrete capacitors [9]. Obviously, stray voltages of this nature are undesirable in an integrated circuit and can cause reliability problems, logic errors, etc., especially if the supply voltages have been scaled down with the device sizes.

The close physical proximity of neighboring lines is not the only cause of increased coupling in ULSI. The increasing length of interconnects with increasing chip size also contributes to increased coupling since a larger region is available for interaction when the interconnects are long. Suffice it to say, then, that crosstalk and coupling can be a serious problem in large ULSI chips because of the dense packing and long interconnect length. To understand the nature of this coupling and finally to find ways of countering it, one needs to be able to describe and model such coupling effects within a basic microscopic formalism. In this paper, we present such a model derived from general coupled mode theory [10]–[18]. The formalism is perfectly general and applies for coupling between both optical and silicide interconnects, i.e., for both optical coupling and quantum mechanical coupling (tunneling). The applicability of the model for both types of coupling is a rather fortuitous coincidence which happens because the basic equations describing the two types of coupling at the microscopic level are the electromagnetic wave equation (derived from Maxwell's equation) and the Schrödinger equation which are mathematically similar. In the next section, we present the theory and then in Section III we estimate coupling coefficient

Manuscript received August 23, 1991; revised December 4, 1991. This work was supported by the Office of Naval Research by Grant N00014-91-J-1505 and by IBM through a Faculty Development Award.

The author is with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556.

IEEE Log Number 9107782.

cients for both types of coupling. In Section IV we present results for GaAs-AlGaAs optical interconnects and silicide interconnects. Finally, in Section V, we present the conclusions.

II. THEORY

A. Coupling Between Optical Interconnects

To model electromagnetic coupling between a set of closely spaced optical interconnect lines, we view and the interconnects as optical waveguides and start from the wave equation that governs the propagation of an electromagnetic signal in a waveguide. We assume that the interconnects are nonlossy and nondispersive. *This is a very good assumption for optical interconnects comprising GaAs waveguides surrounded by AlGaAs cladding.* The scalar wave equation for a TE mode propagating in the \bar{x} direction in any one waveguide reads [10]

$$\nabla^2 E_y(\bar{r}, t) = \mu \epsilon \frac{\partial^2 E_y}{\partial t^2} + \mu \frac{\partial^2}{\partial t^2} [P_{\text{coup}}(\bar{r}, t)]_y \quad (1)$$

where E_y is the y component of the electric field in the interconnect (waveguide) and $[P_{\text{coup}}]_y$ is the y component of a distributed polarization source caused by the coupling of signal from other interconnects. The quantities μ and ϵ are the permeability and permittivity of the interconnects.

To solve for the field E_y in the above equation, we invoke standard coupled mode theory. The solution $E_y(\bar{r}, t)$ can be written as a linear superposition of the normal modes (unperturbed fields) in the individual interconnects

$$E_y(\bar{r}, t) = \sum_n C_n(x) \mathcal{E}_y^{(n)}(y, z) e^{i\omega t} \quad (2)$$

where $\mathcal{E}_y^{(n)}(y, z)$ is the y component of the electric field in the isolated n th interconnect (in the absence of coupling) and ω is the signal frequency. The field $\mathcal{E}_y^{(n)}(y, z)$ satisfies the unperturbed wave equation so that

$$\begin{aligned} \left[\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right] \mathcal{E}_y^{(n)}(y, z) e^{i\omega t} \\ = \mu \epsilon \frac{\partial^2}{\partial t^2} \mathcal{E}_y^{(n)}(y, z) e^{i\omega t} = -\omega_n^2 \mu \epsilon \mathcal{E}_y^{(n)}(y, z) e^{i\omega t} \end{aligned} \quad (3)$$

where ω_n is the signal frequency in the n th interconnect.

Substituting (2) in (1), we get

$$\begin{aligned} \sum_n \left\{ \frac{\partial^2 C_n}{\partial x^2} + C_n \left[\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right] \right\} \mathcal{E}_y^{(n)}(y, z) e^{i\omega t} \\ = -\mu \epsilon \omega^2 \left[\sum_n C_n \mathcal{E}_y^{(n)}(y, z) \right] e^{i\omega t} \\ + \mu \frac{\partial^2}{\partial t^2} [\hat{P}_{\text{coup}}(\bar{r}, t)]_y \left(\sum_n C_n \mathcal{E}_y^{(n)}(y, z) \right) e^{i\omega t} \end{aligned} \quad (4)$$

where $[\hat{P}_{\text{coup}}(\bar{r}, t)]_y$ is an operator such that $[\hat{P}_{\text{coup}}(\bar{r}, t)]_y \mathcal{E}_y = [P_{\text{coup}}(\bar{r}, t)]_y$.

Using (3), we can replace the terms within the square brackets in the left-hand side of (4) to get

$$\begin{aligned} \sum_n \left[\frac{\partial^2 C_n}{\partial x^2} - \mu \epsilon \omega_n^2 C_n \right] \mathcal{E}_y^{(n)} \\ = -\mu \epsilon \omega^2 \sum_n [C_n \mathcal{E}_y^{(n)}] + \mu \frac{\partial^2}{\partial t^2} [\hat{P}_{\text{coup}}]_y \left(\sum_n C_n \mathcal{E}_y^{(n)} \right). \end{aligned} \quad (5)$$

Finally, multiplying the above equation by $\mathcal{E}_y^{(m)*}$ (the asterisk denotes complex conjugate) and integrating over all space, we obtain the coupled set of equations for the coefficients C_n

$$\sum_n \left\{ \frac{\partial^2 C_n}{\partial x^2} O_{mn} + \mu \epsilon (\omega^2 - \omega_n^2) C_n O_{mn} - \kappa_{m,n}^2 C_n \right\} = 0 \quad (m = 1, 2, 3, \dots, n) \quad (6)$$

where $\kappa_{m,n}^2 = \int d^3 r \mathcal{E}_y^{(m)*} \mu (\partial^2 / \partial t^2) [\hat{P}_{\text{coup}}]_y \mathcal{E}_y^{(n)}$ and O_{mn} is the overlap between the fields in the m th and n th interconnect ($O_{mn} = \int d^3 r \mathcal{E}_y^{(m)*} \mathcal{E}_y^{(n)}$).

Equation (6) is a set of n coupled second order differential equations. The difference between (6) and the equations of conventional coupled mode theory is that we have not assumed $O_{mn} = 0$ for $m \neq n$. In fact, this assumption would be incorrect in the limit of strong coupling where the overlap between the fields in neighboring interconnects can be quite significant and neglecting this overlap may result in violation of energy conservation [13]–[18].

Let us now make the following substitution

$$\hat{D}_n(x) = C_n(x) \exp \left(-i \int k_n dx \right) \quad (7)$$

where

$$(k_n)^2 = \mu \epsilon (\omega^2 - \omega_n^2). \quad (8)$$

Substituting (7) in (6) and using (8), we get

$$\begin{aligned} \sum_n \left[\frac{\partial^2 \hat{D}_n}{\partial x^2} O_{mn} + 2ik_n \frac{\partial \hat{D}_n}{\partial x} O_{mn} - \kappa_{mn}^2 \hat{D}_n \right] \\ \cdot \exp \left(i \int k_n dx \right) = 0 \quad (m = 1, 2, 3, \dots, n). \end{aligned} \quad (9)$$

Equation (9) can be recast in a matrix form

$$[\mathbf{A}] \frac{\partial^2}{\partial x^2} [\hat{\mathbf{D}}] + 2i[\mathbf{B}] \frac{\partial}{\partial x} [\hat{\mathbf{D}}] = [\mathbf{K}] [\hat{\mathbf{D}}] \quad (10)$$

where $[\hat{\mathbf{D}}]$ is an $n \times 1$ matrix whose elements are the coefficients $\hat{D}_1, \hat{D}_2, \dots, \hat{D}_n$, $[\mathbf{A}]$ is an $n \times n$ matrix whose elements are $A_{mn} = O_{mn} \exp(i \int k_n dx)$, $[\mathbf{B}]$ is an $n \times n$ matrix whose elements are $B_{mn} = k_n O_{mn} \exp(i \int k_n dx)$ and $[\mathbf{K}]$ is an $n \times n$ matrix whose elements are given by $K_{mn} = \kappa_{mn}^2 \exp(i \int k_n dx)$.

Equation 10 is the general coupled mode equations for a system of n optical interconnects. To illustrate the usefulness of these equations, we now proceed to solve them

for the case of just two interconnects. For this, we first note that ratio of the second to the first term in the LHS of (10) is of the order of the ratio of the distance scale over which significant coupling occurs to the wavelength of the signal in the interconnects. For electromagnetic coupling, the signal wavelength is the wavelength of the optical or electromagnetic signal which is between 1 and 100 μm typically. (We are only concerned with high speed interconnects, i.e., optical interconnects or those that carry ultrashort pulses or millimeter waves.) For quantum mechanical coupling the wavelength is the DeBroglie wavelength of electrons which is between 10 and 100 \AA . Since even for the most densely packed interconnects, we do not expect significant coupling to occur over such small scales, we can always neglect the first term in the LHS of (10) in comparison with the second term. This allows us to obtain closed analytical solutions of (10) in the case of two interconnects.

If we neglect the first term in the LHS (10), we get

$$\begin{aligned}\frac{\partial \hat{D}_1}{\partial x} &= -i\Delta_1 \hat{D}_1 - i\Omega_{12} \hat{D}_2 \exp \left\{ -i \int [k_1 - k_2] dx \right\} \\ \frac{\partial \hat{D}_2}{\partial x} &= -i\Delta_2 \hat{D}_2 - i\Omega_{21} \hat{D}_1 \exp \left\{ -i \int [k_2 - k_1] dx \right\}\end{aligned}\quad (11)$$

where

$$\begin{aligned}\Delta_1 &= \frac{1}{2} \left[\frac{\kappa_{11}^2 O_{22}}{k_1(O_{11}O_{22} - |O_{12}|^2)} - \frac{\kappa_{21}^2 O_{12}}{k_2(O_{11}O_{22} - |O_{12}|^2)} \right] \\ \Delta_2 &= \frac{1}{2} \left[\frac{\kappa_{22}^2 O_{11}}{k_2(O_{11}O_{22} - |O_{12}|^2)} - \frac{\kappa_{12}^2 O_{21}}{k_1(O_{11}O_{22} - |O_{12}|^2)} \right] \\ \Omega_{21} &= \frac{1}{2} \left[\frac{\kappa_{21}^2 O_{11}}{k_2(O_{11}O_{22} - |O_{12}|^2)} - \frac{\kappa_{11}^2 O_{21}}{k_1(O_{11}O_{22} - |O_{12}|^2)} \right] \\ \Omega_{12} &= \frac{1}{2} \left[\frac{\kappa_{12}^2 O_{22}}{k_1(O_{11}O_{22} - |O_{12}|^2)} - \frac{\kappa_{22}^2 O_{12}}{k_2(O_{11}O_{22} - |O_{12}|^2)} \right]\end{aligned}\quad (12)$$

We now make another transformation of variables

$$\begin{aligned}\hat{D}_1 &= D_1 \exp \left(-i \int \Delta_1 dx \right) \\ \hat{D}_2 &= D_2 \exp \left(-i \int \Delta_2 dx \right) \\ k_1 &= k'_1 + \Delta_1 \\ k_2 &= k'_2 + \Delta_2.\end{aligned}\quad (13)$$

This reduces (11) to

$$\begin{aligned}\frac{\partial D_1}{\partial x} &= -i\Omega_{12} D_2 \exp \left\{ -i \int [k'_1 - k'_2] dx \right\} \\ \frac{\partial D_2}{\partial x} &= -i\Omega_{21} D_1 \exp \left\{ -i \int [k'_2 - k'_1] dx \right\}.\end{aligned}\quad (14)$$

The above equations can be decoupled to yield

$$\frac{\partial^2 D_1}{\partial x^2} + i[k'_1 - k'_2] \frac{\partial D_1}{\partial x} + \Omega_{12}\Omega_{21} D_1 = 0 \quad (15)$$

whose solution is

$$D_1(x) = e^{i\delta x} [P e^{i\nu x} + Q e^{-i\nu x}] \quad (16)$$

where P and Q are constants, $2\delta = k'_2 - k'_1$ and $\nu = \sqrt{\kappa^2 + \delta^2}$ where $\kappa = \Omega_{12} = \Omega_{21}$. The last equality follows from the fact that the two interconnects are assumed to be identical.

Assuming that the wave vectors are independent of position x , we get from (7), (13), and (15)

$$C_1(x) = e^{ik_0 x} [P e^{i\nu x} + Q e^{-i\nu x}] \quad (17)$$

where $k_0 = k'_2 + k'_1$.

Similarly, we get

$$C_2(x) = -e^{ik_0 x} [P' e^{i\nu x} + Q' e^{-i\nu x}] \quad (18)$$

where $P' = Pk^+/\kappa$ and $Q' = Qk^-/\kappa$ with $k^+ = 2\delta + 2\nu$ and $k^- = 2\delta - 2\nu$.

To evaluate the constants P and Q , we need to apply the boundary conditions. Let

$$\begin{aligned}C_1(x=0) &= P + Q = A_1^+ \\ C_2(x=0) &= -\frac{k^+ P + k^- Q}{\kappa} = A_2^+\end{aligned}\quad (19)$$

This gives

$$\begin{aligned}P &= -\frac{k^-}{2\nu} A_1^+ - \frac{\kappa}{2\nu} A_2^- \\ Q &= \frac{k^+}{2\nu} A_1^+ + \frac{\kappa}{2\nu} A_2^-\end{aligned}\quad (20)$$

Now let

$$\begin{aligned}C_1(x=L) &= B_1^+ \\ C_2(x=L) &= B_2^+\end{aligned}\quad (21)$$

Using (16)–(20), we get

$$\begin{aligned}B_1^+ &= e^{ik_0 L} \left[\left(\cos(\nu L) + \frac{i\delta}{\nu} \sin(\nu L) \right) A_1^+ \right. \\ &\quad \left. + \left(-\frac{i\kappa}{\nu} \sin(\nu L) \right) A_2^+ \right] \\ B_2^+ &= e^{ik_0 L} \left[\left(\cos(\nu L) - \frac{i\delta}{\nu} \sin(\nu L) \right) A_2^+ \right. \\ &\quad \left. + \left(-\frac{i\kappa}{\nu} \sin(\nu L) \right) A_1^+ \right].\end{aligned}\quad (22)$$

This equation can be written in a matrix form as

$$\begin{pmatrix} B_1^+ \\ B_2^+ \end{pmatrix} = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \cdot \begin{pmatrix} A_1^+ \\ A_2^+ \end{pmatrix} \quad (23)$$

where

$$\begin{aligned} a &= e^{i\kappa L} \left(\cos(\nu L) + \frac{i\delta}{\nu} \sin(\nu L) \right) \\ b &= c = e^{i\kappa L} \left(-\frac{i\kappa}{\nu} \sin(\nu L) \right) \\ d &= e^{i\kappa L} \left(\cos(\nu L) - \frac{i\delta}{\nu} \sin(\nu L) \right). \end{aligned} \quad (24)$$

Note that the 2×2 matrix in (23) is in the form of a transmission matrix. The matrix element b ($= c$) is an indication of the coupling from one interconnect to another. It is the fraction of the signal in interconnect 1 that gets coupled to interconnect 2 after a distance L . Substituting back the values of κ and ν , we find that the quantity b is given by

$$|b| = |c| = \frac{\kappa}{\sqrt{\kappa^2 + \delta^2}} \sin[\sqrt{\kappa^2 + \delta^2} L]. \quad (25)$$

It is easy to show that this coupling is maximum if $\delta = 0$ as long as $\tan(\kappa L) > \kappa L$ or as long as $\kappa L \leq \pi/2$. To have $\delta = 0$ would require that the two interconnects be identical and carry the same signal frequency. Note that when δ is zero it is possible for 100% of the signal in one interconnect to get coupled to the other and this happens at a distance

$$L_{100\%} = \frac{\pi}{2\kappa}. \quad (26)$$

In the case of $\delta = 0$, the coupling over a distance L is simply given by

$$|b| = |c| = |\sin(\kappa L)| \quad (27)$$

Therefore, the fraction of the signal power from one interconnection that is coupled into another is given by

$$|b|^2 = |c|^2 = |\sin(\kappa L)|^2. \quad (28)$$

The most important step now is to derive an expression for the coupling constant κ . But before we proceed to do so, we show that the Schrödinger equation governing quantum mechanical coupling between interconnect lines is mathematically similar to the wave equation governing electromagnetic coupling between optical interconnects so that the Schrödinger equation yields similar solutions for the coupling parameter b ($= c$). In fact, the expression for b (for quantum mechanical tunneling) is identical to the expression given in (27).

B. Quantum Mechanical Coupling

The equations governing quantum mechanical transport of electrons through an interconnect (which can be viewed as an electron waveguide) is the Schrödinger equation which reads

$$i\hbar \frac{\partial \psi}{\partial t} = (H_0 + H')\psi = \left(-\frac{\hbar^2}{2m} \nabla^2 + H' \right) \psi \quad (29)$$

where \hbar is the reduced Planck's constant, m is the electron's effective mass, ψ ($= \psi(\vec{r}, t)$) is the electronic wave

function, H_0 is the unperturbed Hamiltonian, and H' is the perturbation in the Hamiltonian arising from coupling.

We will assume that the perturbation is static (this neglects phonon-assisted or incoherent tunneling) and use time-independent perturbation theory. The wave function is written as

$$\begin{aligned} \psi(\vec{r}, t) &= \psi(x, y, z, t) = [C_1(x)\phi_1(y, z) \\ &+ C_2(x)\phi_2(y, z)]e^{-iEt/\hbar} \end{aligned} \quad (30)$$

where $\phi_{1,2}(y, z)e^{-iE_{1,2}t/\hbar}$ are the unperturbed wave functions in interconnects 1 and 2 in the absence of coupling. These wave functions are eigen functions of H_0 (which is Hermitean) and therefore orthonormal.

Since $\phi_{1,2}(y, z)e^{-iE_{1,2}t/\hbar}$ are eigen functions of H_0 , i.e., they satisfy the unperturbed Schrödinger equation, it is easy to see that

$$H_0\phi_{1,2}(y, z) = E_{1,2}\phi_{1,2}(y, z). \quad (31)$$

Substituting (30) in (29), we get

$$\begin{aligned} &-\frac{\hbar^2}{2m} \left\{ \frac{\partial^2 C_1}{\partial x^2} + C_1 \left[\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right] \right\} \phi_1 \\ &-\frac{\hbar^2}{2m} \left\{ \frac{\partial^2 C_2}{\partial x^2} + C_2 \left[\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right] \right\} \phi_2 \\ &= E[C_1\phi_1 + C_2\phi_2] - H'(C_1\phi_1 + C_2\phi_2) \end{aligned} \quad (32)$$

where $\phi_{1,2} \equiv \phi_{1,2}(y, z)e^{-iE_{1,2}t/\hbar}$.

Comparing the above equation with (5), we find that they are mathematically similar. This is not surprising since both the electromagnetic wave equation and the Schrödinger equation describe the propagation of a wave through a medium. Because of this mathematical similarity we can write down the coupling coefficients $|b|$ and $|c|$ from analogy without any further derivation.

$$|b| = |c| = \sin(\zeta L) \quad (33)$$

where

$$\begin{aligned} \zeta &= \frac{(m|H'_{1,2}|/\hbar^2)O_{11} - (m|H'_{1,1}|/\hbar^2)O_{21}}{\beta(O_{11}O_{22} - |O_{12}|^2)} \\ &= \frac{(m|H'_{2,1}|/\hbar^2)O_{22} - (m|H'_{2,2}|/\hbar^2)O_{12}}{\beta(O_{11}O_{22} - |O_{12}|^2)} \end{aligned} \quad (34)$$

while $H'_{m,n} = \int d^3r \phi_m^* H' \phi_n$ and

$$\beta = \sqrt{2m(E - E_1)/\hbar} = \sqrt{2m(E - E_2)/\hbar} \quad (35)$$

where E_1 and E_2 are the unperturbed energy levels in interconnect 1 and 2.

III. CALCULATION OF COUPLING COEFFICIENTS

A. Optical Coupling Coefficient

To calculate the optical coupling coefficient κ , we consider two identical parallel or crossing interconnects which we view as optical waveguides. The configuration is shown in Fig. 1(a) and (b). We concentrate on TE₁₀ wave propagation and one particular transverse mode. Fig. 1(c) shows the refractive index profile $\eta(z)$ along the z direction where $\eta(z) = \sqrt{\mu_r(z)\epsilon_r(z)}$ and the subscript r denotes

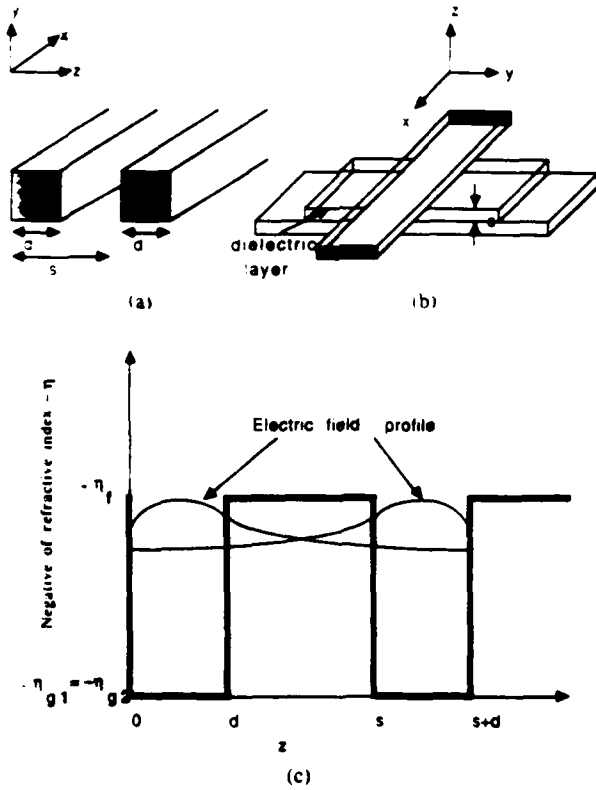


Fig. 1. (a) Two parallel interconnects of width d and separation s . (b) two crossing interconnects, and (c) the refractive index profile along with the electric field distribution in two neighboring optical interconnects that act as waveguides.

relative permeability or permittivity. The refractive indexes in the two interconnects are η_{e1} and η_{e2} , respectively, while the refractive index of the intervening medium is η_f .

We now calculate the quantities κ_{11} ($= \kappa_{22}$) and κ_{12} ($= \kappa_{21}$) which appear in the expressions for $\Omega_{12} = \Omega_{21} = \kappa$.

$$\begin{aligned}\kappa_{11}^2 &= \omega^2 \mu_0 \epsilon_0 \int dz \mathcal{E}_v^{(1)}(z) (\eta_f - \eta_{e1}) \mathcal{E}_v^{(1)}(z) \\ \kappa_{12}^2 &= \omega^2 \mu_0 \epsilon_0 \int dz \mathcal{E}_v^{(1)}(z) (\eta_f - \eta_{e1}) \mathcal{E}_v^{(2)}(z) \quad (36)\end{aligned}$$

$$O_{12} = O_{21}$$

$$\begin{aligned}&= e^{-\alpha_s s} E_g^2 \left[\frac{\eta_g^2 - N^2}{\eta_g^2 - \eta_f^2} \left\{ \frac{1}{\alpha_s} + e^{\alpha_s d} (s - d) \right\} \right] \\&+ e^{-\alpha_s s} E_g^2 \left[\frac{\eta_g^2 - N^2}{\eta_g^2 - \eta_f^2} \left\{ \frac{\alpha_s (e^{\alpha_s d} \cos(\gamma_g d + \phi) - \cos \phi)}{\alpha_s^2 + \gamma_g^2} \right\} \right] \\&+ e^{-\alpha_s s} E_g^2 \left[\frac{\eta_g^2 - N^2}{\eta_g^2 - \eta_f^2} \left\{ \frac{\gamma_g (e^{\alpha_s d} \sin(\gamma_g d + \phi) - \sin \phi)}{\alpha_s^2 + \gamma_g^2} \right\} \right] \\&+ e^{-\alpha_s s} E_g^2 \left[\frac{\eta_g^2 - N^2}{\eta_g^2 - \eta_f^2} \left\{ \frac{\alpha_s (e^{\alpha_s d} \cos \phi - \cos(\gamma_g d + \phi))}{\alpha_s^2 + \gamma_g^2} \right\} \right] \\&+ e^{-\alpha_s s} E_g^2 \left[\frac{\eta_g^2 - N^2}{\eta_g^2 - \eta_f^2} \left\{ \frac{\gamma_g (\sin(\gamma_g d + \phi) - e^{\alpha_s d} \sin \phi)}{\alpha_s^2 + \gamma_g^2} \right\} \right] \quad (41)\end{aligned}$$

where μ_0 and ϵ_0 are the free space permeabilities and permittivities.

We will assume that the two interconnects are identical so that $\eta_{e1} = \eta_{e2} = \eta_e$. The electric field amplitudes are given by [see Fig. 1(c)] [12]

$$\mathcal{E}_v^{(1)}(z) = \begin{cases} E_f e^{\alpha_s z} & z \leq 0 \\ E_g \cos(\gamma_g z + \phi) & 0 \leq z \leq d \\ E_f \exp[-\alpha_s(z - d)] & z \geq d \end{cases} \quad (37)$$

$$\mathcal{E}_v^{(2)}(z) = \begin{cases} E_f \exp[\alpha_s(z - s)] & z \leq s \\ E_g \cos(\gamma_g(z - s) + \phi) & s \leq z \leq s + d \\ E_f \exp[-\alpha_s(z - s - d)] & z \geq s + d \end{cases} \quad (38)$$

where

$$\begin{aligned}\alpha_s^2 &= k^2 - \eta_f^2 \omega^2 \mu_0 \epsilon_0 \\ \gamma_g^2 &= \eta_g^2 \omega^2 \mu_0 \epsilon_0 - k^2 \\ E_g^2 (\eta_g^2 - N^2) &= E_f^2 (\eta_g^2 - \eta_f^2) \\ \phi &= -\tan^{-1} \left(\frac{\alpha_s}{\gamma_g} \right) \\ N &= k / \omega \sqrt{\mu_0 \epsilon_0} \quad (39)\end{aligned}$$

and k is defined in (8).

The quantities O_{11} , O_{22} , O_{12} , κ_{11} , and κ_{12} can now be computed easily.

$$\begin{aligned}O_{11} &= O_{22} \\ &= \frac{E_g^2}{4\gamma_g} \left[4 \frac{\eta_g^2 - N^2}{\eta_g^2 - \eta_f^2} \cot \phi + 2\gamma_g d \right. \\ &\quad \left. + \sin(2\gamma_g d + 2\phi) - \sin 2\phi \right] \quad (40)\end{aligned}$$

$$\kappa_{12}^2 = e^{-2\alpha s} \omega^2 \mu_0 \epsilon_0 \sqrt{\frac{\eta_r^2 - N^2}{\eta_r^2 - \eta_i^2}} E_0^2 \left[\frac{\sin(\gamma_r d)}{\sqrt{\alpha^2 + \gamma_r^2}} \right] \quad (42)$$

$$\kappa_{11}^2 = e^{-2\alpha s} \omega^2 \mu_0 \epsilon_0 \left(\frac{\eta_r^2 - N^2}{\eta_r + \eta_i} \right) \frac{E_0^2}{2\alpha_r} [e^{2\alpha_r d} - 1]. \quad (43)$$

One can now use the results of (40)–(43) in (12) to obtain κ and then use it in (27) to find the coupling parameter b for electromagnetic coupling.

B. Quantum Mechanical Coupling Coefficient

To estimate the coupling coefficient ζ associated with quantum mechanical tunneling between two interconnects, we first refer to Fig. 2 which shows two parallel interconnects [Fig. 2(a)] or two crossing interconnects [Fig. 2(b)] as a system of two coupled quantum wires. Fig. 2(c) plots the potential energy profile $V(z)$ that an electron sees along the z direction. The electron's kinetic energy of motion in the z direction is quantized into subband states that are labeled by n (E_n is the energy of the n th subband state as shown in Fig. 2). The height of the potential barrier is $V - V_1$ for the first interconnect and $V - V_2$ for the second interconnect. The distance between the interconnects is s and the width of each interconnect (dimension along z direction) is d .

To calculate ζ , we first find the quantities $|H'_{1,2}|$, O_{12} , O_{11} , and O_{22} which appear in the expression for ζ [see (34)]. Following [15] we calculate these quantities as follows

$$|H'_{1,2}| = \left| \int_{-\infty}^{\infty} \psi_1^* [V - V_2] \psi_2 dz \right|. \quad (44)$$

The wave functions $\psi_1(z)$ and $\psi_2(z)$ are given by [18]

$$\psi_1(z) = \begin{cases} e^{\alpha z} \cos \theta & z \leq 0 \\ \cos(\gamma z + \theta) & 0 \leq z \leq d \\ \exp[-\alpha(z - d)] \cos(\gamma d + \theta) & z \geq d \end{cases} \quad (45)$$

$$\psi_2(z) = \begin{cases} e^{\alpha(z-s)} \cos \theta & z \leq s \\ \cos(\gamma(z-s) + \theta) & s \leq z \leq s+d \\ \exp[-\alpha(z-s-d)] \cos(\gamma d + \theta) & z \geq s+d \end{cases} \quad (46)$$

where

$$\begin{aligned} \alpha &= \sqrt{2m/\hbar^2 (V - E_n)} \\ \gamma &= \sqrt{2m/\hbar^2 (E_n - V_1)} \\ \theta &= -\tan^{-1}(\alpha/\gamma) \\ C &= \sqrt{\frac{2}{d + 2/\alpha}} \end{aligned} \quad (47)$$

The energy E_n is the energy of an electron in the n th bound state in a quantum well and is found from the eigen equa-

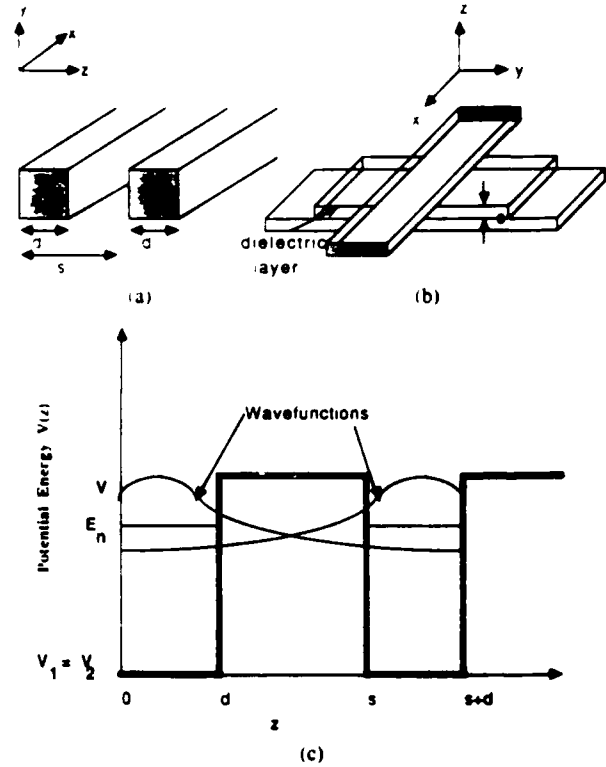


Fig. 2. (a) Two parallel interconnects of width d and separation s , (b) two crossing interconnects, and (c) the potential energy profile along with the wavefunctions in two neighboring narrow interconnects that act as quantum wires.

tion

$$\gamma d = 2 \tan^{-1}(\alpha/\gamma) + n\pi. \quad (48)$$

The quantities O_{11} ($= O_{22}$), O_{12} ($= O_{21}$), $H'_{1,1}$ and $H'_{1,2}$ can now be calculated easily

$$\begin{aligned} O_{11} &= O_{22} = 1 \\ O_{12} &= O_{21} \\ &= \frac{2e^{-\alpha s}}{d + 2/\alpha} \left[\frac{\cos^2 \theta + \cos^2(\gamma d + \theta)}{2\alpha} \right. \\ &\quad \left. + e^{\alpha d} \cos(\gamma d + \theta) \cos \theta(s - d) \right] \\ &\quad + \frac{2e^{-\alpha s}}{d + 2/\alpha} \\ &\quad \cdot \cos \theta \left[\frac{\alpha(e^{\alpha d} \cos(\gamma d + \theta) - \cos \theta)}{\alpha^2 + \gamma^2} \right] \\ &\quad + \frac{2e^{-\alpha s}}{d + 2/\alpha} \cos \theta \left[\frac{\gamma(e^{\alpha d} \sin(\gamma d + \theta) - \sin \theta)}{\alpha^2 + \gamma^2} \right] \\ &\quad + \frac{2e^{-\alpha s}}{d + 2/\alpha} \cos(\gamma d + \theta) \\ &\quad \cdot \left[\frac{\alpha(e^{\alpha d} \cos \theta - \cos(\gamma d + \theta))}{\alpha^2 + \gamma^2} \right] \\ &\quad + \frac{2e^{-\alpha s}}{d + 2/\alpha} \cos(\gamma d + \theta) \\ &\quad \cdot \left[\frac{\gamma(\sin(\gamma d + \theta) - e^{\alpha d} \sin \theta)}{\alpha^2 + \gamma^2} \right] \end{aligned} \quad (50)$$

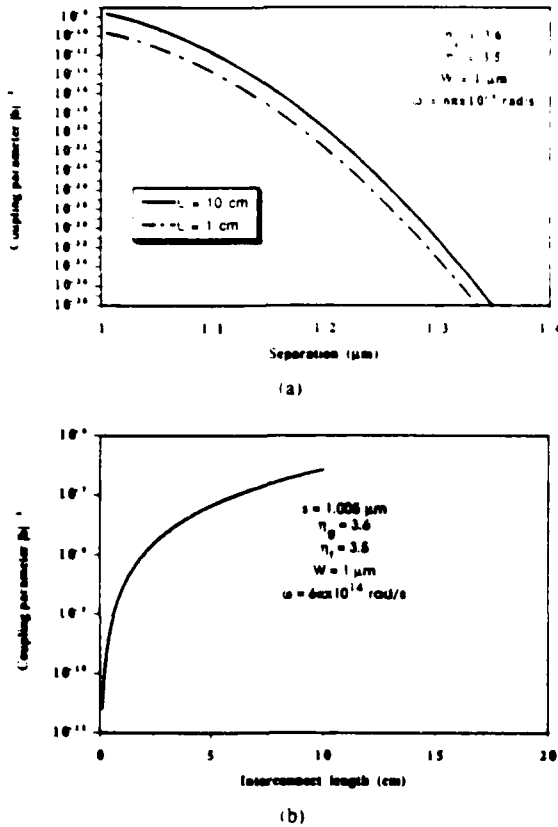


Fig. 3. (a) The coupling parameter $|b|^2$ (which is the fraction of the power in one interconnect coupled to the other) for electromagnetic coupling versus the separation between the interconnects. The interconnects are $1\text{ }\mu\text{m}$ wide and the refractive indexes of the interconnect material and the surroundings are 3.6 and 3.5, respectively. The results are plotted for two different lengths of the interconnects, namely 1 and 10 cm. The angular frequency is $6\pi \times 10^{14}\text{ rad/s}$. (b) The coupling parameter $|b|^2$ versus the length of the interconnect for a spacing of $1.005\text{ }\mu\text{m}$. All other parameters are the same as in Fig. 3(a).

$$H_{12} = e^{-\alpha s} \frac{2(V - V_2) \cos(\gamma d + \theta) \sin(\gamma d)}{d + 2/\alpha \sqrt{\gamma^2 + \alpha^2}} \quad (51)$$

$$|H_{12}| = e^{-2\alpha s} \frac{2(V - V_2) \cos^2(\gamma d + \theta)}{d + 2/\alpha \cdot 2\alpha} [e^{2\alpha d} - 1]. \quad (52)$$

We can use the results of (49)–(52) in (33)–(35) to obtain the coupling parameter b in the case of quantum mechanical coupling.

IV. RESULTS

In Fig. 3(a) we show the coupling parameter $|b|^2$ due to optical coupling as a function of the spacing s between two identical optical interconnects for two different lengths of the interconnects. The coupling parameter $|b|^2$, as mentioned before, is the fraction of the power in one interconnect that is coupled to the other. The interconnects are assumed to be $1\text{ }\mu\text{m}$ wide (this is of the order of the wavelength of light emitted by semiconductor lasers and light emitting diodes), the material is GaAs with a refractive index of 3.6 and the isolating medium is Al-

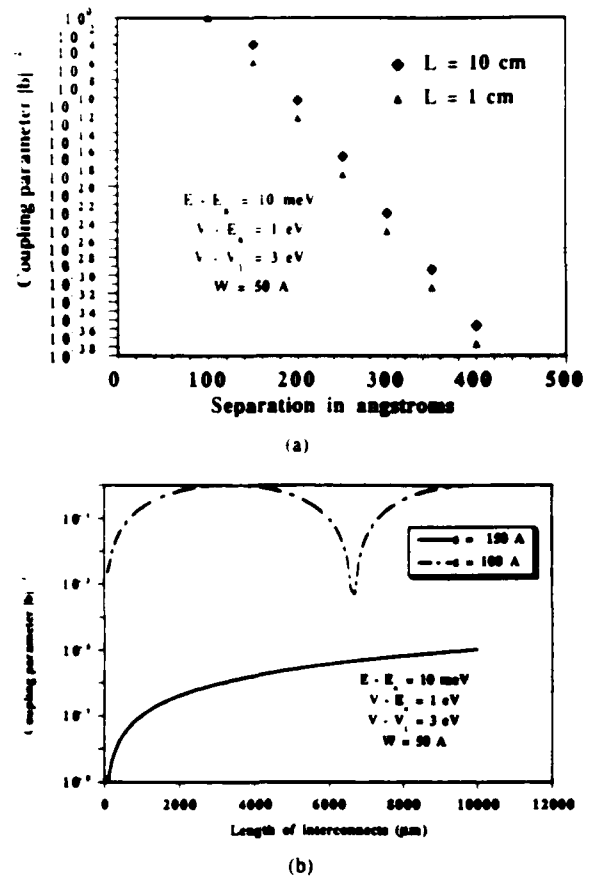


Fig. 4. (a) The coupling parameter $|b|^2$ for quantum mechanical coupling versus the separation between interconnects. The interconnects are assumed to be made of polysilicon embedded in silicon dioxide. The energy barrier between polysilicon and oxide is assumed to be 3 eV, the electron subband energy is 1 eV below the barrier and the longitudinal kinetic energy is 10 meV. The width W of the interconnects is $50\text{ }\text{\AA}$. The results are plotted for two different lengths of the interconnects, namely 1 and 10 cm. (b) The coupling parameter $|b|^2$ versus the length of the interconnect for two different spacings of 100 and $150\text{ }\text{\AA}$. All other parameters are the same as in Fig. 4(a).

GaAs with a refractive index of 3.5. Fig. 3(b) shows the coupling parameter due to electromagnetic coupling as a function of interconnect length when the spacing between the optical interconnects is $1.005\text{ }\mu\text{m}$. The signal frequency is assumed to be $2\pi \times 10^{14}\text{ Hz}$ which roughly corresponds to the signal frequency of a GaAs diode laser. Other parameters regarding the interconnects are shown in figure legends. In Fig. 4(a) we show the coupling parameter due to quantum mechanical coupling as a function of spacing between two identical silicide interconnects embedded in silicon dioxide. We assume that the potential barrier between silicide and silicon dioxide is 3 eV which is close to the potential barrier between silicon and silicon dioxide. The interconnects are assumed to be $50\text{ }\text{\AA}$ wide (the limit of present-day lithographic capability). The data is presented for two different lengths L of the interconnects ($L = 1\text{ cm}$ and $L = 10\text{ cm}$). In Fig. 4(b), we show the coupling parameter as a function of the length of the interconnect for two different spacings of 100 and $150\text{ }\text{\AA}$. All other relevant parameters about the interconnects are displaced in the legend.

V. CONCLUSION

It is evident from the results that coupling in optical interconnects is not a serious problem in ULSI. This is because GaAs-AlGaAs waveguides provide excellent confinement of the optical signal. However, quantum mechanical tunneling can be quite serious in silicide dielectrics embedded in silicon dioxide. As shown in Fig. 4(a), the coupling parameter $|b|^2$ for quantum mechanical coupling approaches 1 if the spacing is 100 Å. The coupling is always larger for a longer length of the interconnects since a larger region for coupling is provided. As shown in Fig. 4(b), the coupling is also larger when the spacing between interconnects is smaller. In Fig. 4(b), we see that the coupling oscillates with increasing length of the interconnect L when the spacing is small enough. This is easily understood from (33) which shows that $|b|^2 = \sin^2(\zeta L)$ which is an oscillatory function in L . When the spacing is 100 Å, the coupling and ζ are large enough that a period of the oscillation occurs when $L = 6300$ Å. For a spacing of 150 Å, the coupling and ζ are much smaller so that a much larger length L would be required for a full period of the oscillation. Therefore, we do not see the oscillation for the spacing of 150 Å in Fig. 4(b).

Fig. 4 shows the quantum mechanical tunneling between silicide interconnects can be a serious problem in ULSI. Tunneling, of course, can be a coherent process or an incoherent process (phonon-assisted tunneling, for instance, is incoherent) and our analysis has dealt with coherent tunneling only. In our analysis, we assumed that the electron wavefunction is coherent over the entire width of the interconnect which is a good assumption for very narrow polysilicon interconnects (<50 Å wide) even at room temperature and certainly below. Consequently, tunneling may cause significant crosstalk in ULSI circuits at room temperature and at cryogenic temperatures. Since many ultrafast electronic devices such as high electron mobility transistors (HEMT's) or quantum coupled devices utilizing resonant tunneling transistors may be constrained to operate at liquid nitrogen temperatures (77 K), quantum mechanical tunneling between interconnects could pose a serious problem at these temperatures. To circumvent this problem, one may devise ways of destroying the coherence of the electron wave function. One possibility is to impregnate the intervening dielectric isolating the interconnects with a soft magnetic material with strong spin-orbit scattering. Since spin orbit scattering is very efficient in destroying coherence, this may inhibit significant tunneling between neighboring interconnects and reduce crosstalk. The other obvious way of counteracting this problem is to use as isolating dielectrics those insulators that present a large energy barrier to tunneling. In this respect, silicon dioxide is better than silicon nitride since the energy barrier to tunneling is usually larger with silicon dioxide which has an energy gap of ~9 eV compared to ~5 eV for silicon nitride. The small energy gap of silicon nitride may preclude its use as isolating dielectric in ULSI in spite of its other attractive properties such as high resistance to Na⁺ diffusion.

In conclusion, we have presented a model to calculate crosstalk between optical interconnects using waveguides and between narrow and closely spaced silicide interconnects embedded in dielectrics. We find that the crosstalk between silicide interconnects arising from quantum mechanical tunneling could be serious in ULSI circuits. We have also suggested some remedies for that problem.

REFERENCES

- [1] H. G. Craighead, "Nanolithography for ultrasmall structure fabrication," in *Physics of Quantum Electron Devices*, F. Capasso, Ed. New York: Springer-Verlag, 1989.
- [2] H. I. Smith, K. Ismail, W. Chu, A. Yen, Y. C. Ku, M. L. Schattenburg, and D. A. Antoniadis, "Fabrication of quantum-effect electronic devices using x-ray nanolithography," in *Nanostructure Physics and Fabrication*, ed. M. A. Reed and W. P. Kirk, Eds. Boston, MA: Academic, 1989.
- [3] E. E. Ehrichs, and A. L. deLozanne, "Fabrication of nanometer features with a scanning tunneling microscope," in *Nanostructure Physics and Fabrication*, M. A. Reed and W. P. Kirk, Eds. Boston, MA: Academic, 1989.
- [4] S. P. Beaumont, "Fabrication and overgrowth of quantum wires and dots for optoelectronic applications," in *Nanostructure Physics and Fabrication*, M. A. Reed and W. P. Kirk, Eds. Boston, MA: Academic, 1989.
- [5] —, "Nanofabrication and optical assessment of quantum wires and dots," in *Granular Nanoelectronics*, D. K. Ferry, J. R. Barker, and C. Jacobini, Ed., NATO ASI Series. New York: Plenum, 1991.
- [6] D. K. Ferry, "Lateral surface superlattices and the future of ULSI microelectronics," in *Granular Nanoelectronics*, D. K. Ferry, J. R. Barker, and C. Jacobini, Eds. NATO ASI Series. New York: Plenum, 1991.
- [7] J. M. Mikkelsen, L. A. Hall, A. K. Malhotra, S. D. Seccombe, and M. S. Wilson, *IEEE J. Solid State Circ.*, vol. 16, p. 542, 1981.
- [8] S. I. Long and S. E. Butler, *GaAs Digital Integrated Circuit Design*. New York: McGraw Hill, 1990.
- [9] T. A. Fulton and G. J. Dolan, "Observation of single electron charging effects in small tunnel junctions," *Phys. Rev. Lett.*, vol. 59, p. 109, 1987.
- [10] A. Yariv, "Coupled mode theory for guided wave optics," *IEEE J. Quantum Electron.*, vol. QE-9, pp. 919-933, 1973.
- [11] A. Ishimaru, *Electromagnetic Wave Propagation, Radiation and Scattering*. Englewood Cliffs, NJ: Prentice Hall, 1991.
- [12] T. Tamir, Ed., *Integrated Optics*. New York: Springer Verlag, 1975.
- [13] A. Hardy and W. Streifer, "Coupled mode theory of parallel waveguides," *J. Lightwave Technol.*, vol. LT-3, p. 1135, 1985.
- [14] —, "Coupled mode solutions of multiwaveguide systems," *IEEE J. Quantum Electron.*, vol. QE-22, p. 528, 1986.
- [15] L. Tsang and S. L. Chuang, "Improved coupled mode theory for reciprocal anisotropic waveguides," *J. Lightwave Technol.*, vol. LT-6, p. 304, 1988.
- [16] S. L. Chuang, "A coupled mode formulation by reciprocity and a variational principle," *J. Lightwave Technol.*, vol. LT-5, p. 5, 1987.
- [17] —, "A coupled mode theory for multiwaveguide systems satisfying reciprocity," *J. Lightwave Technol.*, vol. LT-5, p. 174, 1987.
- [18] S. L. Chuang and B. Do, "Electron states in two coupled quantum wells—A strong coupling of modes approach," *J. Appl. Phys.*, vol. 62, p. 1290, 1987.

Supriyo Bandyopadhyay was born in Calcutta, India. He received the B.Tech. in electronics and electrical communication engineering from the Indian Institute of Technology, Kharagpur, the M.S. degree in electrical engineering from Southern Illinois University, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN.

From 1986 to 1987 he was a Visiting Assistant Professor at Purdue University. He joined the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN, in 1987 where he is now an Associate Professor. His research interests are in quantum and hot electron transport and cryogenic measurements.

New high-contrast developers for poly(methyl methacrylate) resist

Gary H. Bernstein^{a)}

Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

Davide A. Hill

Department of Chemical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

Wen-Ping Liu^{b)}

Center for Solid State Electronics Research, Arizona State University, Tempe, Arizona 85287-6206

(Received 3 October 1991; accepted for publication 8 January 1992)

New developers for poly(methyl methacrylate) consisting of mixtures of common developing components have been carefully investigated. It has been found that adding a small percentage of methyl ethyl ketone to methyl isobutyl ketone and Cellosolve results in a significant increase in contrast. Results of contrast experiments as well as improvements in electron-beam lithographic exposures are reported. An explanation of the mechanism of contrast and resolution enhancement is offered.

1. INTRODUCTION

Very large scale integration (VLSI) technology continues to push toward smaller geometries with advancement in the $\frac{1}{2}$ μm range and experimental circuits such as monolithic microwave integrated circuits (MMICs) in the 0.1 μm range. The fabrication of nanostructures and quantum devices below 0.1 μm especially relies on ultrahigh lithographic resolution, and as such continues to utilize primarily poly(methyl methacrylate) (PMMA) as the resist of choice. PMMA continues to find applications in electron-beam lithography (EBL), focused ion-beam lithography, x-ray lithography, deep uv optical lithography, and, more recently, scanning tunneling lithography.¹ It is almost universally the choice for the development of new lithographic techniques since PMMA offers the highest known resolution of any organic resist. A drawback to the use of PMMA is that it is generally not sensitive enough for manufacturing purposes, but it does offer extremely high resolution. For this reason, PMMA remains technologically important.

The generally accepted criterion of resolution prediction of a resist-developer system has fallen on contrast γ . High contrast can be thought of as the ability of a developer to discriminate between very small differences in exposure dose from one small area to an adjoining one. Exposure in electron-beam lithography is not precisely localized due to factors including the Gaussian beam cross section, forward scattering in the resist, backscattering from the resist and substrate (proximity effect), and the generation of secondary electrons in the resist.² These factors decrease the modulation transfer function of the exposure for closely spaced patterns.

In PMMA, a positive resist, bonds are broken by incident radiation (e.g., an electron beam), decreasing the average molecular weight M of the long-chain molecules. This increases the solubility of the resist in a suitable de-

veloper allowing the exposed pattern to be indented into the resist. In further processing steps, the pattern is transferred by various techniques to the substrate material. An important limitation on the density of the exposed pattern, and in some respects the individual line-widths, is the proximity effect which causes exposure of the resist up to several micrometers away from the area of the primary exposure. As lines are placed closer together, the total dose increases between the primary exposed regions, thus leading to a decreased differential in exposure.² The requirements of the developer to distinguish between differences in dose become more stringent if the pattern is to be resolved.

Ideally, a developer will remove all positive resist that has been exposed to a dose above a certain threshold and none of the resist that has been exposed below that threshold. In practice, a range of exposures results in partial development of the resist over that range. For this reason, higher contrasts allow a narrower area to be developed completely even in the presence of those factors that broaden the exposure. For other lithography schemes mentioned above, similar energy scattering limitations to resolution exist. It is expected that for all types of lithography, higher resolution will result from higher contrast, and the improvements discussed below will be equally relevant to any application of PMMA.

Contrast is a measure of the sharpness of the developing threshold of the developer-resist system and is defined as the slope of the plot of normalized thickness of the resist remaining after development versus the logarithm of the exposure dose. Figure 1 illustrates the definition of γ . The onset of development D_i is defined to be the dose at the intersection of the steepest part of the contrast curve extrapolated to the 100% level of the curve. The sensitivity D_f is defined to be the dose of the extrapolated curve at zero thickness.

As discussed above, γ is a measure of the ability of the developer to distinguish between gradations of dose. Since contrast is expressed as

$$\gamma = [\log(D_i/D_f)]^{-1}. \quad (1)$$

^{a)} Author to whom correspondence should be addressed.

^{b)} Present Address: VLSI Research Department, Lishan Microelectronics Co., P. O. Box 19, Lintong, Shaanxi 710600, Peoples Republic of China.

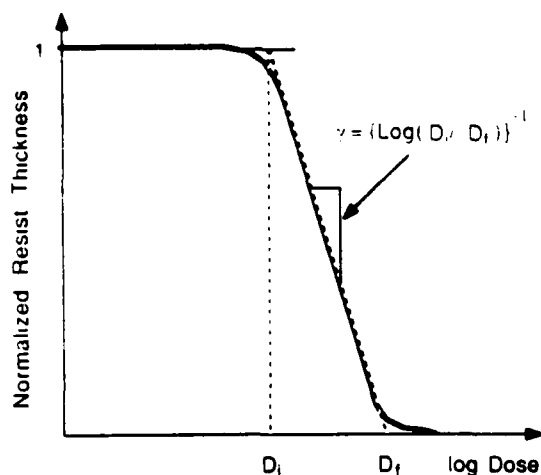


FIG. 1. Plot of normalized positive resist thickness remaining after development as a function of log of exposure dose (the contrast curve) illustrating the definitions of D_i and D_f .

a higher absolute value of γ represents a higher contrast developer. (In this report, the absolute value of all contrast values will be used, but are all negative.) The ideal situation discussed above in which there is an absolute dose threshold for dissolution represents an infinite contrast, whereas in practice the contrast is always finite. Contrast is commonly measured experimentally for a given development time by varying the exposure dose in small increments through the region of dose that causes the resist to dissolve until total dissolution is achieved. A common method for measuring γ is to use a surface profilometer to measure the thickness of resist remaining after exposure and development and plot the data as discussed above.³⁻⁹

As we discuss later, the contrast parameter is entirely dictated by the dissolution behavior of the polymer. Papanu *et al.* have produced valuable fundamental work on the mechanism of PMMA dissolution in one- and two-solvent developers, and for different polymer molecular weights.¹⁰ Their experiments greatly clarify the role of thermodynamic quality of the solvent on the rate and mechanism of polymer dissolution, but only for one- and two-component developer mixtures. Here we approach the more complex case of a multicomponent developer and, on more practical grounds, we study the effect of solvent quality (in our case various mixtures) on the rate of dissolution of electron-beam-exposed PMMA resists. We explore combinations of common developers to ascertain whether or not improvements in the lithographic process could be promoted simply by using readily available mixtures of solvents.

These new developer solutions are mixtures of chemicals that have been commonly employed individually as developers for PMMA, but which when combined exhibit higher performance than when alone. Our combination of these chemicals has shown improvements in achievable contrast of PMMA as well as improved lithographic re-

sults. Some of our developer solutions give improvements of up to 35% in contrast and no loss in sensitivity.

Three chemical solutions are often employed as PMMA developers. These are (i) methyl ethyl ketone:ethanol (EtOH) in a ratio of 26.5:73.5; (ii) Cellosolve (2-ethoxyethanol):methanol (MeOH) in a ratio of 3:7; and (iii) methyl isobutyl ketone:2-propanol (IPA) in a ratio of 1:3. These solutions are referred to here simply as solution A, or S-A, solution B, or S-B, and solution C, or S-C, respectively. Pure methyl ethyl ketone, Cellosolve, and methyl isobutyl ketone will be represented by MEK, CS, and MIBK, respectively. S-C has been reported to provide very high resolution at exposures of 50 kV electrons with a development time of 15–45 s^{11,12} while S-B has been reported to provide very high resolution in EBL exposures between 20 and 120 kV with a development time of 5 s.^{13,14} It is mixtures of these three common developer solutions which we have found to give markedly improved results in contrast.

II. EXPERIMENT

For resist exposures, we used an ISI-100B scanning electron microscope (SEM) modified for EBL in a vector-scanned mode with an exposure spot (pixel) spacing and beam diameter of 100 nm each.¹⁵ All test exposures were performed at 40 kV on thick silicon substrates. All PMMA films were of molecular weight 950 000 a.u., spin coated and baked at 160 °C for 4 h, with resist thicknesses ranging from 0.4 to 1.0 μm . Exposures were arranged in a 6×6 array of $50 \times 50 \mu\text{m}^2$ squares. Resist thicknesses remaining after development were obtained using a Dektak II surface profilometer.

S-A, S-B, and S-C were tested alone and in combination using development times that yielded approximately the same dose for complete development and also corresponded to typical values from the literature. These times were 45 s for S-C,^{11,12} 5 s for S-B,^{13,14} and 2 s for S-A. The development time for almost all mixtures of these components was 10 s. This time was based upon the relative strength and concentrations of the components used in the mixtures and was estimated initially to yield the same time to complete development for approximately similar doses. The range of developer temperatures represents common fluctuations caused by varying room conditions, evaporative cooling, etc.

Development was performed by holding the samples with tweezers and dipping in the developer with light agitation for the required time. Temperatures during development were determined by inserting a mercury-bulb thermometer, calibrated to 1 °C, into the beaker. Developer times were controlled as much as possible by careful attention to the process. Significant error occurred only in the possible case of S-A with a 2 s develop (estimated at less than 20%), but this data was used only for crude comparison as will be discussed below. All samples were rinsed for 15 s in MeOH immediately after development and blown dry in nitrogen.

The unexposed dissolution properties of the three developer components were tested to ensure that normalized

TABLE I. Results of contrast experiment identified by sample number, concentrations, and development temperature.

Sample no.	Volume Percent				Develop time (s)	Temperature °C	γ
	S-C	S-B	S-A	MEK			
1	0	100	0	0	5	19	5.7
2	0	100	0	0	5	21	7.0
3	100	0	0	0	45	19	7.0
4	100	0	0	0	45	21	9.3
5	66.4	32.0	1.6	0.424	10	---	7.7
6	66.4	32.0	1.6	0.424	10	---	7.0
7	49.6	48.0	2.4	0.636	10	---	8.2
8	49.6	48.0	2.4	0.636	10	---	8.5
9	50.1	45.4	4.5	1.19	10	18	7.9
10	50.1	45.4	4.5	1.19	10	18	7.9
11	50.1	45.4	4.5	1.19	10	19	9.4
12	50.1	45.4	4.5	1.19	10	19	9.4
13	50.1	45.4	4.5	1.19	10	19	9.1
14	50.1	45.4	4.5	1.19	10	19	8.8
15	0	95	5	1.19	5	21	12.8
16	0	0	100	100	2	19	4.7

thicknesses were not compromised by excessive dissolution rates. This was necessary since the profilometer technique can only measure resulting resist thicknesses with respect to the unexposed portions. By soaking for several hours, S-B and S-C were found to yield unexposed dissolution rates of 2.8 and 0.5 Å/min, respectively. The harshest developer, S-A, displayed the highest dissolution rate of 51 Å/s. However, even this rate was still small enough to be negligible given our initial thickness and short develop times for mixtures containing even small amounts of MEK.

Data were taken for the three developer components alone as well as for a variety of mixtures of the three components. Contrast was determined from the steepest slope on the graphs of percent resist remaining versus the log of exposure dose. The steepest slopes were determined using a linear regression routine applied to the three steepest data points. Special care was taken to apply exactly the same criterion for the determination of steepest slope in all cases.

The mixtures of the developers were investigated to show the effects of varying primarily the S-A concentration. It was deemed that S-A is such a harsh developer that its overall concentration should be very low, and also that the properties of S-B and S-C are similar enough that their concentrations should also be similar. However, higher concentrations of S-C were included in the mixtures since it yields generally higher contrast than S-B. Since S-A was found to give low contrast by itself at an MEK concentration of only 26.5%, we investigated primarily the range of very small MEK concentrations. For these reasons, the three mixtures investigated were S-C:S-B:S-A in ratios of 42:20:1, 21:20:1, and 11:10:1. We consider the most important parameter to be the S-A concentration, which in the above mixtures is 1.59%, 2.38%, and 4.55%, respectively. This corresponds to an absolute concentration of MEK in the total mixture of only 0.42%, 0.64%, and 1.2%, respectively. Please note that mixture ratios are expressed in units of S-A volumes, and percentages are provided in Table I.

III. RESULTS AND DISCUSSION

The results of our contrast experiment are shown in Figs. 2 and 3. Figure 2 shows contrast curves for the three original developers, S-A, S-C, and S-B. For the data shown, the contrast for S-A is 4.7, for S-C 7.0, and for S-B 5.7. It is important to note, as will be demonstrated below, that small differences in contrast can be very important to lithographic results, given the logarithmic nature of the plot.

Figure 3 shows contrast curves for two mixtures of the three common solutions. It can be seen by comparing Figs. 2 and 3 that the sensitivity of the mixtures is approximately the same as that of S-C, falling midway between that of S-A and S-B. This is an extremely desirable trait of a developer, namely, that an increase in contrast does not lead to a decrease in sensitivity of the resist-developer system. This is more significant considering that the develop time for the mixtures is only 10 s compared with 45 s for S-C.

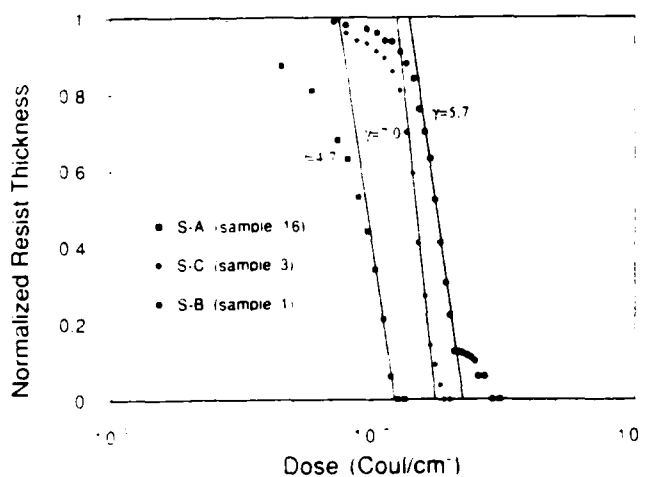


FIG. 2. Contrast curves for the three original developers.

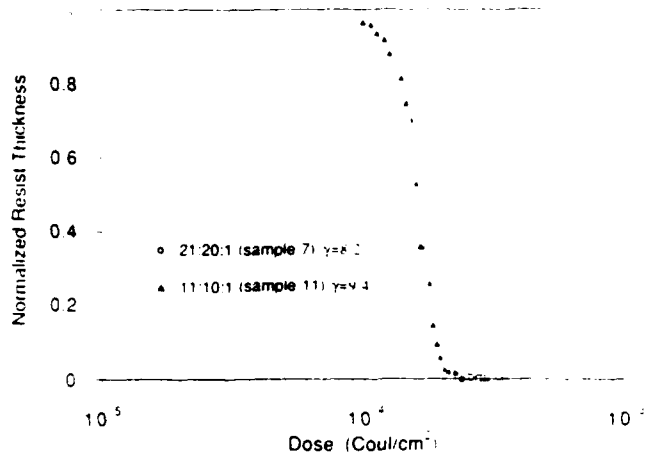


FIG. 3. Contrast curves for mixtures of the original developer components. Ratios are S-C:S-B:S-A.

This implies that the mixture must offer higher sensitivity than S-C or S-B alone if development times are increased. In addition, it is shown in Appendix A that contrast is predicted to increase for increased develop times but only for an increased solubility, as in our case of added MEK. This implies that further study would indicate a possibility of achieving even higher contrasts coupled with improved sensitivity for longer develop times. In practice, the optimum exposures for highest resolution correspond to the dose at D_f of Fig. 1, where D_f must decrease for increased develop time. This has strong implications for the viability of improving the throughput of processes utilizing PMMA for high-resolution commercial applications.

We have observed markedly improved lithographic performance using the new developer mixtures. Figure 4 shows photomicrographs of test patterns consisting of sets of parallel lines of decreasing pitch to $0.1 \mu\text{m}$ in the direction of the outside of the pattern, and overlapping to form a variable-spaced grid in the corner. Both test patterns were exposed on a single sample under the identical conditions of resist thickness, bake temperature, bake time, dose, line spacing, and beam energy. After exposure the patterns were separated and developed as discussed above using S-C for pattern 4(a) and S-C:S-B:S-A 11:10:1 for pattern 4(b).

The resulting improved contrast can be observed from the variations in light and dark areas. One expects that as the parallel lines are exposed closer together, the proximity effects cause some exposure of the spaces between the lines.² A lower-contrast developer will dissolve the PMMA at the lines of exposure and also remove much of the resist between the lines. This is the cause of the light areas in Fig. 4(a) as the line pitch decreases. However, a sufficiently higher-contrast developer will remove mostly the resist in areas of direct exposure and less of the resist between the lines. This is the case in Fig. 4(b) where there is clearly less light area, showing that the resist remains intact between the lines even in the presence of identical proximity effects to part 4(a). These results cannot be explained as a

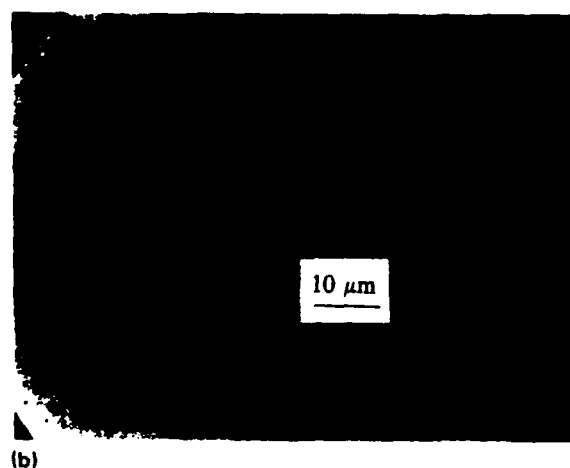
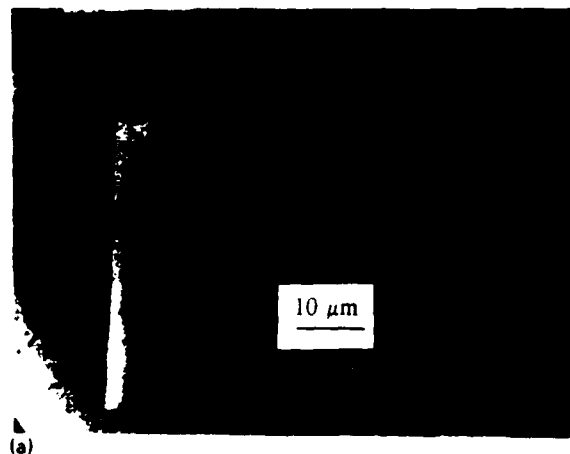
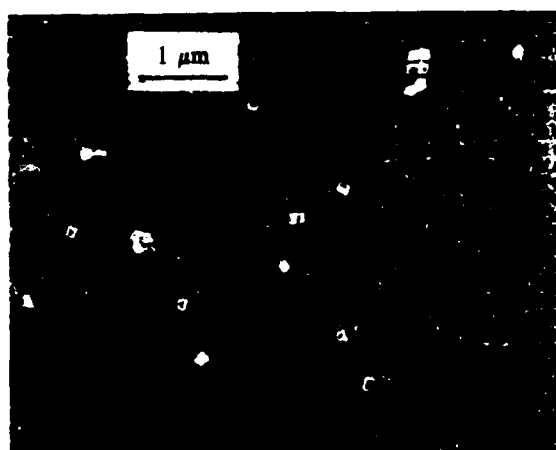


FIG. 4. Photomicrographs of test patterns illustrating improved contrast from new developers. The test patterns consist of parallel lines of decreasing pitch in the outward directions, crossing in a variable-pitch grid pattern at the corner. Light areas indicate severe loss of resist between the exposed lines. Both patterns were exposed under identical conditions, but developed by (a) S-C for 45 s and (b) S-C:S-B:S-A 11:10:1 for 10 s. The appearance of less light area in (b) implies that the mixture has provided much higher contrast than that of S-C in (a).

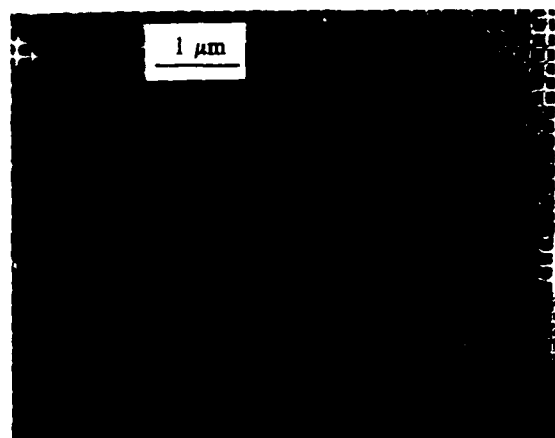
simple matter of sensitivity and overdeveloped exposure, since the sensitivity of the 11:10:1 mixture is almost exactly the same as that of S-C as can be seen by comparing Figs. 2 and 3.

Figure 5(a) shows a grid pattern developed by the S-B solution. The linewidths vary by about 25% toward the interior of the pattern. Figure 5(b) shows a pattern developed in S-C:S-B:S-A 11:10:1. There is almost no evidence of linewidth variations over the pattern due to the higher contrast of the 11:10:1 mixture versus that of the S-B. In addition, lift-off properties are affected by contrast. Figure 5(a) shows a typical example of metal flakes left from the unexposed areas after a lift-off step in acetone. Figure 5(b), however, shows the consistently improved lift-off achieved due to steeper sidewalls resulting from the higher contrast. The patterns of Figs. 4 and 5, then, help demonstrate that the new mixtures provide useably higher contrast compared with conventional developers.

Table I shows the contrast for various percentages of developer solutions and temperatures. A plot of contrast as



(a)



(b)

FIG. 5. Grid patterns formed by EBL exposure at 40 kV on GaAs substrates followed by development and lift-off of 40 nm gold: (a) developed in S-B and (b) developed in S-C:S-B:S-A 11:10:1. Note decreased line broadening in (b) vs (a) indicating higher contrast of the mixture over S-B. Also note typical improved lift-off as a result of steeper edge profiles due to higher contrast.

a function of percent S-A, uncorrected for temperature dependence, is shown in Fig. 6. A trend of increasing γ as S-A is increased is indicated. The same data, but normalized to 19 °C, are shown in Fig. 7. (Note that contrast data shows no correlation with the initial resist thickness, which is accounted for in the process of normalizing the remaining thickness after development.) The temperature normalization assumes an Arrhenius behavior with an activation energy of 24.3 kcal/mol, found by fitting contrast data at a fixed S-A percentage as a function of temperature. This value is intermediate between that found by Cooper, Krasicky, and Rodriguez¹⁶ of 25 kcal/mol for MEK and that reported by Greeneich¹⁷ of 24 kcal/mol for MIBK. In the case of Fig. 7, a trend of increasing contrast for increasing percent MEK is now clearly evident. The data, including S-C, collapse to a straight line of the least-squares fit very closely except for the case of S-B at 0% MEK. (Precise temperatures were not available for the case of 0.42% and 0.64% MEK. The bar in the figure is not an error bar, but represents calculated uncertainty of plus or minus 0.45 °C based on the known value of the activation energy centered about 19 °C.) It is striking that within this linear region,

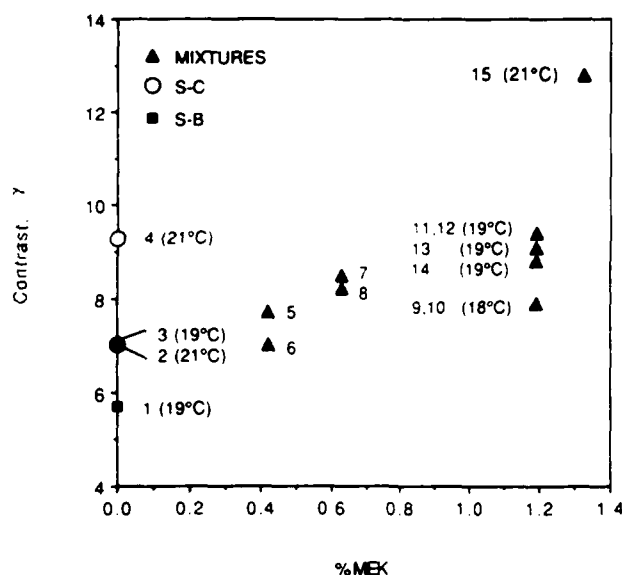


FIG. 6. Plot of γ as a function of percent MEK, uncorrected for temperature dependence. Sample number and temperature are noted for each datum.

about a 1.3% increase of MEK concentration in the mixture results in more than a 35% increase in contrast, and that further improvements can be promoted by simply increasing the temperature.

The fact that S-B does not fall on the temperature-compensated line can be qualitatively understood by noting

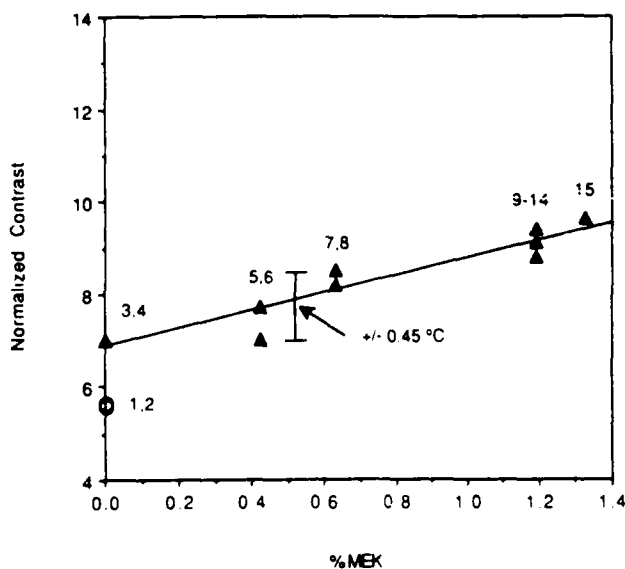


FIG. 7. Plot of γ as a function of percent MEK, corrected for temperature dependence by normalizing to 19 °C. The normalization assumes an Arrhenius dependence with an activation energy of 24.3 kcal/mol. Notice that the points in Fig. 6 corresponding to sample no. 4 (21 °C) for pure S-C and that of samples no. 9 and 10 (18 °C) at 1.19% MEK have collapsed onto the straight line and are now obscured on the graph by other data points. The bar in the figure is not an error bar, but represents calculated uncertainty of ± 0.45 °C based on the known value of the activation energy centered about 19 °C.

that both MEK and MIBK are ketones, while CS (2-ethoxyethanol) contains both alcohol and ether groups. The chemical compositions of MIBK and MEK are very similar, differing only by two aliphatic carbons. In order to understand the dynamics completely, one would need to consider a full thermodynamic and transport analysis involving six components. A complete analysis is beyond the scope of this work, but a simple explanation of the role of the components is given below.

Previous analyses of the development of EBL-exposed resist show that measured decrements of normalized resist thickness upon exposure (i.e., the contrast curve) should be directly proportional to the rate of polymer dissolution¹⁷ (see also Appendix A). For a given developer system, this rate depends on the average molecular weight of the exposed polymer, which in turn is a function of the dose. Appendix A shows that contrast is expected to exhibit the same temperature behavior as that of dissolution rate and explains our observation of increased γ for developments performed at slightly higher temperatures where dissolution rates are in general higher. We found that at all S-A concentrations, γ increased with temperature between 18 and 21 °C. With the exception of only the S-B data, all other data properly collapsed onto the best-fit line as a function of temperature as suggested by the above theory.

The main factors that affect the rate of solubilization of a polymer in a solvent are the plasticization and thermodynamic compatibility, or solvent "quality." Plasticization refers to the ability of the solvent to penetrate the polymer and increase the free volume,^{18,19} making it more vulnerable to dissolution. The rate at which a solvent can plasticize a polymer is related to its molecular size and therefore its molecular weight. Gipstein *et al.*²⁰ reported a noticeable solvent-size effect in the dissolution rate of atactic PMMA in a homologous series of *n*-alkyl acetates. They found a precipitous drop in dissolution rate for increased penetrant molecular weight. As an approximation we will use molecular weights as an estimator of the relative ability of the solvent components to plasticize the exposed PMMA.

The solvent quality is related to the match between three thermodynamic parameters of the solvent and the polymer. The closer these are, the higher is the thermodynamic compatibility resulting in a higher dissolution rate.¹⁹ These parameters are the three components of the global solubility parameter δ , namely the dispersive, polar, and hydrogen-bonding components, δ_d , δ_p , and δ_h , respectively, where

$$\delta_t^2 = \delta_p^2 + \delta_h^2 + \delta_d^2 \quad (2)$$

We have found it useful to compare all of these components in an attempt to explain our results. In comparing the similarity of the three solubility parameters between the developers and PMMA, we have defined the parameter

$$\Delta = [(\delta_{pi} - \delta_{pj})^2 + (\delta_{hi} - \delta_{hj})^2 + (\delta_{di} - \delta_{dj})^2]^{1/2}, \quad (3)$$

where *i* indicates a developer component and *j* the resist. Since δ_t is the length of the solubility vector in solubility

TABLE II. Molecular weight and solubility parameters for each component of the mixtures. The distance Δ is included for each component.

Component	M (g/mol)	δ_d [(MPa) ^{1/2}]	δ_p [(MPa) ^{1/2}]	δ_h [(MPa) ^{1/2}]	Δ [(MPa) ^{1/2}]
MEK	72	14.0	9.3	9.5	1.4
EtOH	46	12.6	11.2	20.0	12
MIBK	100	14.4	8.1	5.9	3.4
IPA	60	14.0	9.8	16.0	7.5
CS	90	13.0	9.1	15.2	6.8
MeOH	32	11.6	13.0	24.0	16
MMA	90	13.5	10.1	8.5	---

parameter space, Δ is the distance between the end points of these vectors. We therefore refer to Δ as simply "distance."

It is expected that the greater the distance Δ is, then the less efficient the component is at dissolving PMMA, and therefore the lower is the solvent quality. Such measure of compatibility is consistent with the idea that polar, dispersive, and hydrogen-bonding forces can, in principle, act independently. (For example, two chemical species, one strongly polar but nonhydrogen bonding, and the other weakly polar but strongly hydrogen bonding, could have the same global solubility parameter and yet be immiscible. This is the same as saying that the length of δ_t in solubility space is the same for each species, but their directions vary greatly.) Thus only those solubility parameter components pertaining to congruent energetic interactions should be compared.

The calculated values of Δ for all pure solvents compared with MMA, along with the molecular weights of each component, are given in Table II. Again, the plasticization, or molecular weight, of the individual components must also be considered in explaining their role in the solutions. Table III indicates the order of each developer component in its ranking of plasticization and solvent quality. We can see that MEK is indeed the best solvent but rates fourth as a plasticizer. In S-A, the EtOH serves as a diluent and to help in plasticization, being ranked second in this regard. These two facts help to explain why S-A is such an effective solvent of PMMA at 3060 Å/min for a molecular weight as high as 950 000 a.u. This also helps to explain why S-A exhibits the worst contrast, since it strongly dissolves PMMA of all molecular weights. Although CS ranks third in solvent quality, it appears that the presence of MeOH in S-B, ranked first for plasticization, helps to increase the effectiveness of CS as a solvent. Regarding S-C dissolution, the component MIBK is rated

TABLE III. Ranking of plasticization and solvent quality for each component of the developer solutions.

	MEK	EtOH	MIBK	IPA	CS	MeOH
Distance						
Δ [(MPa) ^{1/2}]	1.4	12	3.4	7.5	6.8	16
M (g/mol)	72	46	100	60	90	32
Solvent quality (rank)	1	5	2	4	3	6
Plasticization (rank)	4	2	6	3	5	1

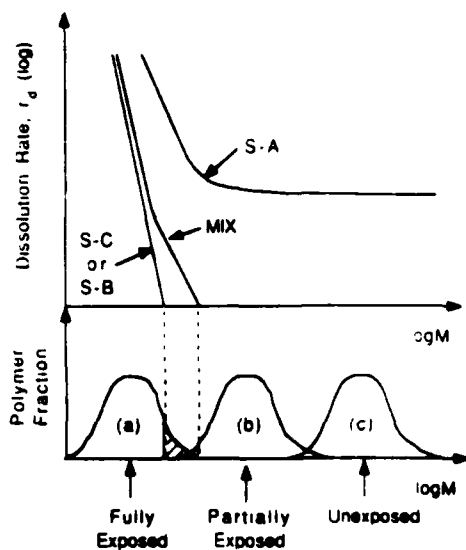


FIG. 8. Explanation of mechanism by which trace amounts of S-A in S-C increases contrast. The solubility curve for S-C alone is compared to the expected curve after the addition of a small amount of S-A (MIX) and S-A alone (see also Ref. 17). Contrast is improved through the increase of r_d for higher molecular weights with only slight change to r_d at low ones. Below the graph are molecular weight distributions for (a) fully exposed, (b) partially exposed, and (c) unexposed PMMA. The shaded region corresponds to those molecular weights that require large exposure doses to be further fragmented. It is preferable to use the lowest exposure dose possible to achieve the highest resolution. The addition of traces of S-A increases the solubility of these fragments.

second as a solvent but is the poorest plasticizer. Isopropanol being ranked only third as a plasticizer helps to explain why S-C dissolves the unexposed PMMA at only 0.5 $\sqrt{\text{min}}$.

As mentioned above, prediction of the solubility properties of systems containing many solvents can be very complex, and only a simple qualitative explanation of those properties can be offered here. Within the exposed resist, a distribution of molecular weights exists whose average depends on the exposure dose.¹² Under conditions of low-level exposure, the inability of a relatively high-contrast developer, such as S-C, to completely remove the PMMA in the primary exposed area is due to the presence of high-molecular-weight polymers that have not been exposed sufficiently to be broken into sufficiently small strands. The presence of a small percentage of MEK increases the overall thermodynamic quality of the solvent, causing those remaining long-chain PMMA strands to be more selectively removed without a major change in the solubility rate of the unexposed or less exposed areas. This argument is supported in Appendix B.

The above argument is plausible when discussed in light of Figure 5 of Ref. 17, which gives solubility rates of MIBK solutions as a function of fragment molecular weight M . We discuss here a two-component system that can be extended to more components. Figure 8 shows a qualitative curve of dissolution rate r_d versus fragment molecular weight M after exposure for a single developer, e.g., S-C, and also a proposed curve for the two-developer mixture S-C + S-A (MIX). Since the dissolution rate for pure

MEK is much higher than for either S-C or dilute MEK in solution, and given the lower value of γ for S-A, r_d for S-A is expected to extend to higher molecular weight with a lower slope and approach a constant for very high molecular weights (similar to MIBK in Ref. 17). Portrayed below in the figure are qualitative molecular weight distributions of a fully, partially, and unexposed polymer. By adding small amounts of MEK to S-C, we have affected the solubility properties of the S-C as shown in the MIX curve to dissolve the higher-molecular-weight fragments far more quickly while changing the slope of the curve only slightly for the lower molecular weight. To first order, we expect the solubility properties to be affected little for lower molecular weights, but increased substantially for higher. (A thermodynamic model that qualitatively supports this assertion is presented in Appendix B where we analyze a two-component developer system by assuming that the rate enhancement derives exclusively from an increase in polymer solubility.) Highest benefit to contrast and resolution is achieved when the cutoff of solubility for the new mixture is above the highest molecular weight of the fragments in the primary exposed area but is below that of the marginally exposed resist for a slightly lower effective dose in a closely adjoining area. This can never be achieved totally in practice, as the distribution of fragment molecular weights in the adjoining areas is a continuum and overlaps that of the polymer in the primary exposed area.

In our case of mixing the three components S-C, S-B, and S-A, Fig. 7 indicates that the contrast improvement is dominated by the presence of MEK. We believe, however, that a more complete study would indicate that each component enhances the solubility in a range of molecular weights yielding an effective characteristic in which all fragmented molecular weights in the primary exposed areas are removed, but little is removed outside. It must be kept in mind that the average molecular weights of the PMMA fragments increase quickly with distance away from the beam. The goal in choosing a correct exposure for achieving high resolution is that the top of the range of fragment molecular weights in the primary exposed area lies just in the range dissolved at an appreciable rate by the trace amounts of MEK in the mixture, so that high-molecular-weight fragments are selectively removed from the area of primary exposure. It is possible to conceptualize an ideal developer whose properties are described in Fig. 9. This developer would have "break points" in the r_d vs M curve that tailor the dissolution rate to the molecular weight distribution curve. This ideal developer would then be assured of dissolving no more of the higher-molecular-weight fragments than was absolutely necessary for a given exposure dose, thereby minimizing the erosion of the lower-molecular-weight fragments (relative to the unexposed areas) in the less-exposed adjoining area. (These arguments are further supported by the theoretical predictions for a binary developer as discussed in Appendix B, and in Fig. 10.) Our three component system is a crude, but demonstrably effective, attempt at achieving this goal.

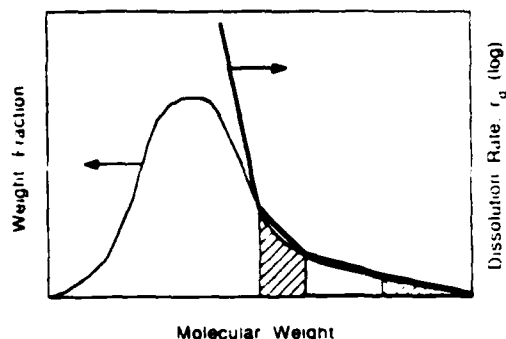


FIG. 9. Solubility curve for a conceptualized "ideal" developer superimposed on the distribution of fragment molecular weights for PMMA. The shape of the solubility curve ensures that dissolution rates for each molecular weight are no higher than necessary to dissolve all molecular weight fractions completely, thereby minimizing erosion of areas adjoining those of primary exposure.

IV. SUMMARY AND CONCLUSIONS

We have investigated new developer mixtures for PMMA positive resist. We demonstrated that adding a small percentage of methyl ethyl ketone to methyl isobutyl ketone/Cellosolve solutions increased contrast substantially. This was illustrated in practice by a comparison of EBL exposures each demonstrating decreased proximity effect.

The mechanism for increased contrast by adding trace amounts of MEK was explained in terms of selective broadening of the range of solubilities for S-A in S-C/S-B solutions. It was concluded that the presence of MEK increased the range of dissolution of fragmented molecular

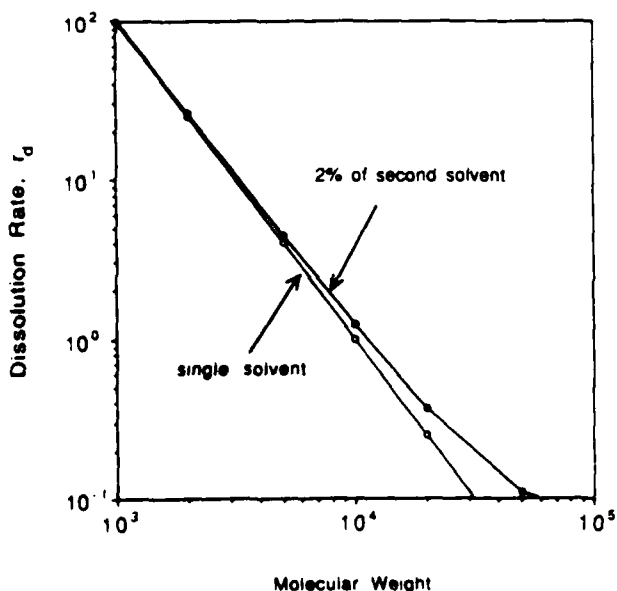


FIG. 10. Calculated rate of dissolution for a single solvent and also that solvent combined with 2% of a second solvent as a function of molecular weight. The mixed curve has been calculated from Eqs. (B7) and (B8) assuming $\beta(x_1/x_2) = M/100$.

weights in the primary exposed regions to more efficiently clear the exposed PMMA without affecting unexposed areas.

We feel that in view of the substantial improvements of the contrast of PMMA and demonstrated improvements in minimizing the proximity effect, the use of PMMA can be further justified for ultrasmall geometries in future specialized integrated circuits and optical and x-ray mask fabrication applications. In addition, as future x-ray sources increase in brightness, the feasibility of using PMMA will become more favorable with the trade-off in support of resolution versus speed becoming more advantageous.

ACKNOWLEDGMENTS

The authors would like to thank J. P. Kohn and D. K. Ferry for helpful discussions. Also, we are grateful to KTI Chemicals for providing materials. This research was supported in part by IBM, AFOSR, ONR, and the University of Notre Dame.

APPENDIX A

Here, we show that the contrast γ is proportional to the rate of polymer dissolution. We consider the dissolution of a polymer film of uniform initial thickness h_0 . The polymer has an average molecular weight M . Exposure to a developer solution for a time Δt_d will cause partial erosion of the film, and the removal of a layer of thickness Δh . In general, we may write for $\Delta h < h_0$,

$$\Delta h/h_0 = r_d \Delta t_d / h_0, \quad (\text{A1})$$

where r_d is the average rate of dissolution, which depends on M , temperature T , and solvent quality. Experiments^{16,17} indicate that

$$r_d = f(M, T, \text{solvent quality}) \exp(-E_d/RT) M^{-\alpha}, \quad (\text{A2})$$

where E_d is the activation energy for dissolution, R the ideal-gas constant, and f a function of solvent quality, polymer molecular weight, and T . The exponent α , a positive real number, may also be a function of M , T , and the solvent. Within certain ranges of M , however, such dependencies of f and α are weak, so that the main effects of T and M are accounted for predominantly by the Arrhenius and power law, respectively.

For a polymer exposed to radiation, the average molecular weight depends on the exposure dose D , i.e. $M = M(D)$. Thus the after-development normalized thickness of an exposed resist will depend on the average polymer molecular weight M , which, in turn, is a function of exposure dose D . The contrast γ , defined as

$$|\gamma| = \max \left(\left| \frac{\partial(\Delta h/h_0)}{\partial \log D} \right| \right), \quad (\text{A3})$$

where the term on the right-hand side can be expressed as

$$\frac{\partial(\Delta h/h_0)}{\partial \log D} = \frac{\partial(r_d \Delta t_d / h_0)}{\partial \log D} = r_d \left(\frac{\partial \log r_d}{\partial \log M} \right) \left(\frac{d \log M}{d \log D} \right) \frac{\Delta t_d}{h_0}. \quad (\text{A4})$$

Equation (A4) shows explicitly a proportionality between γ and r_d . Furthermore, for exclusively Arrhenius temperature dependence and a power-law molecular weight dependence of r_d , as in Eq. (A2), we will have

$$\left(\frac{\partial \log r_d}{\partial \log M}\right) = -\alpha$$

and

$$\left(\frac{d \log M}{d \log D}\right) = \text{function of } D \text{ only,} \quad (\text{A5})$$

which are both independent of temperature. In this case, the contrast will show an Arrhenius temperature dependence, with the same activation energy as the dissolution rate. This supports our observations of increased γ for higher developer temperatures.

Interestingly, from Eq. (A4) we see that for a fixed rate, temperature, and $M(D)$ functionality, a lower value of α will always give lower contrast. Greeneich¹⁷ has shown experimentally for dissolution of PMMA in MIBK:IPA that the slopes of $\log(r_d)$ vs $\log(M)$ decrease appreciably with increasing solvent quality (i.e., for larger MIBK:IPA ratios); indeed, for pure MIBK, after a relatively sharp transition in slope at M around 10^4 a.u., the curves eventually approach a constant (i.e., $\alpha \approx 0$). In this case, according to Eqs. (A4) and (A5) the contrast should either approach zero, or, more realistically, become very small. This is consistent with our finding that films developed in pure S-A, an extremely strong (good) solvent, always showed poor contrast, in fact, poorer than those developed in either S-C or S-B.

APPENDIX B

We propose here a simple explanation of our findings, based on the hypothesis that our observed enhancement of contrast with traces of MEK is due to purely thermodynamic factors. This hypothesis is reasonable, since MEK is such a strong solvent. We assume the ratio of the dissolution rate (and thus contrast) of the system with MEK to that without it to be proportional to the corresponding ratio of polymer volume fractions. This functionality is consistent with the idea that any multiplicative factor quantifying rate enhancement should increase with polymer solubility, and approach unity at zero MEK concentration.

Our task, then, is to determine the increase in polymer volume fraction in solution caused by the presence of traces of a third component in an otherwise binary (polymer + solvent) and biphasic (solid polymer + solution) system. This can be done within the framework of solution thermodynamics.

We consider two equilibrium systems. The first one, a two-phase two-component system, comprises a pure, solid polymer (phase S) in equilibrium with a solution of polymer (component 1) in a solvent (component 2, phase L). The second system (denoted here by an asterisk) is identical to the first one, except for the presence of traces of a second solvent (component 3) in the liquid. We assume

that the polymer is monodisperse, and that the solid precipitate is always made exclusively of polymer, with no appreciable amounts of either solvents.

Thermodynamic equilibrium requires that the chemical potential μ of the polymer in the solution and in the solid must be equal, i.e.,

$$\mu_1^S = \mu_1^L \quad \text{and} \quad \mu_1^{S^*} = \mu_1^{L^*}, \quad (\text{B1})$$

where, again, component 1 is the polymer, and S and L stand for solid and liquid, respectively.

In both cases, the solid is a pure polymer, i.e.,

$$\mu_1^S = \mu_1^{S^*}, \quad (\text{B2})$$

thus

$$\mu_1^L = \mu_1^{L^*}. \quad (\text{B3})$$

Expressions for the chemical potentials of the various components in a ternary mixture of a polymer with two solvents have been given by Flory;²¹ for the polymer we have

$$\begin{aligned} \mu_1 - \mu_1^0 = RT[& \ln \phi_1 + (1 - \phi_1) - \phi_2(x_1/x_2) - \phi_3(x_1/x_3) \\ & + (\chi_{12}\phi_2 + \chi_{13}\phi_3)(\phi_2 + \phi_3) \\ & - \chi_{23}(x_1/x_3)\phi_2\phi_3], \end{aligned} \quad (\text{B4})$$

where the superscript 0 refers to the pure, amorphous, polymer at the same temperature T and pressure of the solution, R is the ideal-gas constant, and x_1 , x_2 , and x_3 are quantities proportional to the molecular volumes of the polymer 1, and solvents 2 and 3, respectively. Additionally, ϕ_i is the volume fraction of species i ($i = 1, 2, 3$), and the interaction parameters χ_{ij} are given by

$$\chi_{ij} = z\Delta w_{ij}x_i/RT, \quad (\text{B5})$$

where z is a lattice coordination number and Δw_{ij} the molar interaction energy between species i and j .

Note that Eqs. (B4) and (B5) have been proven to be of only limited accuracy, even for simple binary systems, and especially for dilute polymer solutions.²² The predicted trends are generally correct, however, and consistent with most experimental evidence. Thus, this model should be able to capture, at least qualitatively, the main physics of our problem.

Equations (B4) and (B5) can be used to calculate the chemical potential of the polymer both in the presence and absence of component 3. Because component 3 is very dilute, the following approximations can be made:

$$|\phi_1^{L^*} - \phi_1^L| \ll 1, \quad \phi_2^L \approx \phi_2^{L^*}, \quad \phi_3^L = 0, \quad \phi_3^{L^*} \ll 1. \quad (\text{B6})$$

From Eqs. (B3)–(B5) we obtain

$$\begin{aligned} \ln(\phi_1^{L^*}/\phi_1^L) = & \phi_3^{L^*}(x_1/x_3)[1 + \phi_2^{L^*}(zx_3/RT) \\ & \times (\Delta w_{23} - \Delta w_{12} - \Delta w_{13})]. \end{aligned} \quad (\text{B7})$$

Thus, an exponential enhancement in polymer solubility is predicted to occur upon addition of traces of component 3, provided that:

$$\beta = [1 + \phi_1^L (zx_2/RT)(\Delta w_{23} - \Delta w_{12} - \Delta w_{13})] > 0. \quad (\text{B8})$$

For $\beta \sim 1$, the increase can be substantial, since (x_1/x_3) is proportional to the ratio of the molecular weight of the polymer to that of component 3. Notice that an exponential increase of the ratio (ϕ_1^L/ϕ_1^L) with polymer molecular weight does not necessarily imply an unbounded increase of polymer volume fraction, since for higher, insoluble, molecular weights, the decay to zero of ϕ_1^L will be faster than any concomitant growth of the enhancement factor. Flory, in his analysis of polymer fractionation by selective precipitation from binary mixtures, derived an equation [see Eq. (23) of Ref. 21] that is very similar to Eq. (B7).

In order to obtain the sign of β in Eq. (B8) it is necessary to estimate the differences of the interaction energies between the various components. Following Gordon and Taylor²³ (see also Ref. 18) in a semiempirical approach, we assume the following proportionality [see Eq. (3)]:

$$\Delta w_{ij} \propto \Delta_{ij}^2 = [(\delta_{pi} - \delta_{pj})^2 + (\delta_{di} - \delta_{dj})^2 + (\delta_{hi} - \delta_{hj})^2], \quad (\text{B9})$$

where i and j denote the particular substances, and δ_p , δ_d , and δ_h are the polar, dispersive, and hydrogen-bonding components of the solubility parameter, respectively. In our case

$$(\Delta w_{23} - \Delta w_{12} - \Delta w_{13}) \propto (\Delta_{23}^2 - \Delta_{12}^2 - \Delta_{13}^2). \quad (\text{B10})$$

Hence, the sign of the interaction energy term on the left-hand side can be estimated from the properties of the pure components. (Further theoretical refinements would require that the factor of proportionality between Δw_{ij} and Δ_{ij}^2 be a function of the molar volumes of the various components. Such refinements are not deemed necessary here, where only rough estimates are of interest.) For a positive right-hand side in Eq. (B10), β in Eq. (B8) will also be positive, indicating an enhancement of polymer solubility. On the other hand, a negative term in Eq. (B10) may either signify an increase or a decrease of solubility, depending on the relative magnitude of the various terms in Eq. (B8).

It is interesting to estimate the sign and magnitude of the quantity in Eq. (B10) for PMMA in MIBK with traces of MEK. This system should be reasonably close to the one we achieved experimentally. Using the pure-component solubility parameters in Table II (MMA = 1, MIBK = 2, MEK = 3), we obtain, in units of cal/ml,

$$\Delta_{23}^2 = 14.49, \quad \Delta_{12}^2 = 11.57, \quad \Delta_{13}^2 = 2,$$

giving

$$\Delta_{23}^2 - \Delta_{12}^2 - \Delta_{13}^2 = 0.92 \text{ cal/ml}.$$

We see that the overall interaction energy is positive, so that we would indeed expect an enhancement of solubility in our case, consistent with observations.

Furthermore, it is possible to extract from our data with Eq. (B7) an "experimental" value of $\beta(x_1/x_3)$. The linear increase of contrast with percent MEK in Fig. 7 is, in fact, consistent with Eq. (B7), since for small values of

the argument, the exponential can be approximated as a straight line. From the temperature-normalized contrast data we obtain $\beta(x_1/x_3) = 24$, which is reasonable, in light of our many approximations, and, possibly, some more fundamental limitations inherent to the thermodynamic model itself. Again, it is worth noticing here that, given the crudeness of our model, such "theoretical" results should be considered only for their qualitative value.

A rate-enhancement curve, calculated from Eq. (B7) by assuming $\phi_3 = 0.02$, $\beta(x_1/x_3) = (M)/100$, and an inverse-square dependence of the dissolution rate on M , is given in Fig. 10 (Note that we have arbitrarily truncated the enhanced-rate curve, since it would have shown an unrealistic upturn for higher molecular weights, due to our assumption of a fixed inverse-square dependence of the original rate for all molecular weights. In reality, due to the finite solubility of higher fractions, the dissolution rate eventually drops to zero much faster than M^{-2} , and no upturn in the net rate occurs). Figure 10 should be compared with curves 1 and 3 in Fig. 5 of Ref. 17.

¹M. A. McCord and R. F. W. Pease, *J. Vac. Sci. Technol. B*, **6**, 293 (1988).

²D. C. Joy, *Microelectron. Eng.*, **1**, 103 (1983).

³W. J. Cooper, P. D. Krasicky, and F. Rodriguez, *Polymer*, **26**, 1069 (1985).

⁴W. M. Moreau, *Opt. Eng.*, **22**, 181 (1983).

⁵S. Mackie and S. P. Beaumont, *Solid State Technol.*, **28**, 117 (1985).

⁶B. K. Daniels, P. Trefonas, and J. C. Woodbrey, *Solid State Technol.*, **31**, 105 (1988).

⁷M. A. Mohsin and J. M. G. Cowie, *Polymer*, **29**, 2130 (1988).

⁸K. G. Chiong, M. B. Rothwell, S. Wind, J. Bucchignano, and F. J. Hohn, *J. Vac. Sci. Technol. B*, **7**, 1771 (1989).

⁹V. K. Sharma, R. Petrick, and S. Afrossman, *Polymer*, **23**, 1732 (1982).

¹⁰J. S. Papanu, D. W. Hess, D. S. Soane, and A. T. Bell, *J. Electrochem. Soc.*, **136**, 3077 (1989).

¹¹A. N. Broers, J. M. E. Harper, and W. W. Molzen, *Appl. Phys. Lett.*, **33**, 392 (1978).

¹²S. P. Beaumont, P. G. Bower, T. Tamamura, and C. D. W. Wilkinson, *Appl. Phys. Lett.*, **38**, 436 (1981).

¹³H. G. Craighead, R. E. Howard, L. D. Jackel, and P. M. Mankiewich, *Appl. Phys. Lett.*, **42**, 38 (1983).

¹⁴A. Muray and M. Isaacson, *J. Vac. Sci. Technol. B*, **1**, 1091 (1983).

¹⁵All exposures were performed at the Center for Solid State Electronics Research at Arizona State University.

¹⁶W. J. Cooper, P. D. Krasicky, and F. Rodriguez, *J. Appl. Polym. Sci.*, **31**, 65 (1986).

¹⁷J. S. Greeneich, *J. Electrochem. Soc. Solid State Sci. Technol.*, **122**, 970 (1975).

¹⁸C. A. Kumins and T. K. Kwei, in *Diffusion in Polymers*, edited by J. Crank and G. S. Park (Academic, New York, 1968), p. 107.

¹⁹F. Rodriguez, *Principles of Polymer Systems*, 3rd ed. (Hemisphere, New York, 1989), p. 36.

²⁰E. Gipstein, A. C. Ouano, D. E. Johnson, and O. U. Need III, *Polymer Eng. Sci.*, **17**, 396 (1977).

²¹P. J. Flory, *Principles of Polymer Chemistry* (Cornell University Press, Ithaca, NY, 1953), p. 549.

²²P. J. Flory, *Principles of Polymer Chemistry* (Cornell University Press, Ithaca, NY, 1953), p. 559.

²³J. Gordon and J. S. Taylor, *J. Appl. Chem.*, **2**, 493 (1952).

Fabrication of ultrasmall tunnel junctions by electron-beam lithography

S. J. Koester,¹⁾ G. Bazán, G. H. Bernstein, and W. Porod

Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

(Received 3 October 1991; accepted for publication 5 November 1991)

Fabrication of a variety of ultrasmall tunneling structures is presented for possible applications in single electronics. Two main structures are described: (i) arrays of nanometer-scale metal dots and (ii) single and multiple linear tunnel junctions formed by overlapping lines. Very large, uniform particle arrays were fabricated by electron-beam lithography (EBL) with particle diameters as small as 25 nm. The advantage of our dot array system is its high degree of uniformity and very small junction size coupled with the use of isolated particles to reduce stray capacitance effects. Single and multiple linear tunnel junctions were fabricated by EBL using single-layer resist and angled evaporations. Our fabrication technique differs from previous ones in its ability to create very small tunnel junctions without the need for multilayer resist systems or precise knowledge of the angle of evaporation.

I. INTRODUCTION

The semiclassical theory of single-electron tunneling (SET) has proven to be useful in describing the behavior of tunnel junctions with ultrasmall capacitances.¹ In this theory, the Coulomb effects due to the individual electrons become increasingly evident as the single-electron charging energy of a structure with capacitance C surpasses the energy of thermal fluctuations at temperature T , i.e.,

$$e^2/2C > kT. \quad (1)$$

Here, e and k are the electron charge and Boltzmann constant, respectively. According to (1), the observation of individual electrons requires cryogenic temperatures, even for the smallest tunnel junctions reported.² This issue has become increasingly important for the future of new devices which might utilize single-electron charging effects.³⁻⁵ Such devices might operate at temperatures above 4.2 K if the characteristic capacitance of a junction can be reduced below 10^{-16} F while maintaining narrow tunnel barriers. Therefore, a major issue facing the growing field of "single electronics" today is the ability to fabricate small, reliable capacitor structures for operation at higher temperatures.

The most common type of system used in studying SET effects is the metal-insulator-metal tunnel junction structure. With this type of system, researchers have utilized two basic approaches to achieving capacitances small enough to experimentally observe single-electron charging. The first technique is to use natural particle arrays as the center electrode of a double-tunnel junction structure.⁶⁻⁸ It is well documented that the particle islands formed by evaporating metal onto an insulator such as Al_2O_3 can have effective capacitances as low as 5×10^{-18} F.⁸ Since the particles are isolated they do not suffer from stray capacitance effects. However, because the particles have widely varying sizes, the temperature at which the SET

effects wash out remains near liquid-helium temperatures. The other technique, pioneered by Dolan,^{9,10} is the use of electron-beam lithography and angled evaporations to create narrow, overlapping metal lines separated by thin tunnel barriers. This technique has proven successful at creating junctions as small as $30 \times 30 \text{ nm}^2$.² However, even very small junctions produced this way can produce large capacitances if the leads connecting the junction significantly couple to their environment or to each other.

The purpose of this work is to investigate improved designs and fabrication techniques which can further enhance the performance of single-electron tunneling structures. This paper focuses only on fabrication issues, while the electrical characterization of these tunneling junctions will be reserved for future publications. In Sec. II, we describe the first of two experimental structures, arrays of ultrasmall metallic dots, along with the fabrication procedure and fabrication results. In Sec. III, we discuss similar issues for systems of linear tunnel junctions formed by overlapping metal lines. Finally, we discuss in Sec. IV some of the pertinent issues involved in creating such structures, as well as the advantages and disadvantages of our approach over those previously reported in the literature.

II. ARRAYS OF ULTRASMALL METAL DOTS

The first structure discussed here is composed of a large array of very small metal "dots" sandwiched between two electrodes and separated by thin tunnel barriers. A schematic of these dot arrays is shown in Fig. 1. Our structure is different from previous ones utilizing isolated particles because the particle array is fabricated by direct exposure of high-resolution electron-beam lithography. The advantage of using EBL to create the particle array is that it can produce uniform dots of equal size and spacing, while still providing very high resolution.

The fabrication procedure for our structures is as follows. The base-electrode metal is evaporated onto a Si substrate through a mechanical shadow mask forming a 10-

¹⁾Present address: Department of Electrical Engineering, University of California at Santa Barbara, Santa Barbara, CA 93106.

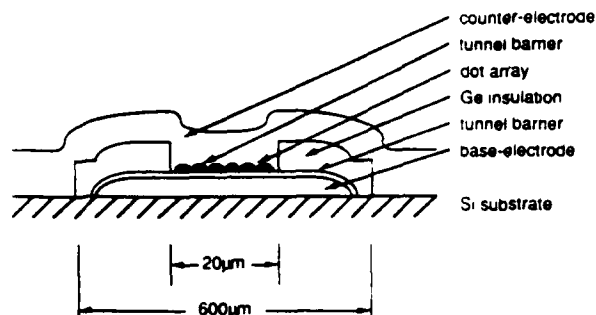


FIG. 1. Cross-sectional diagram of a single-electron tunneling structure utilizing a lithographically fabricated particle array.

mil-wide line. Typically, this metal is Al and the barrier, always an oxide in our case, is formed by exposure to air during the subsequent processing steps. Next, an insulation region is patterned by standard EBL and lift-off to reduce the effective junction area of the structure. This is accomplished by creating a $500 \times 600\text{-}\mu\text{m}^2$ rectangular germanium region with a $20 \times 20\text{-}\mu\text{m}^2$ "hole" in the middle, which covers the base electrode except for the small central area. Next, a $24 \times 24\text{-}\mu\text{m}^2$ dot array is aligned and exposed by EBL to the previous pattern in such a way that the array entirely fills the opening in the Ge insulation. Once again this pattern is evaporated and lifted off, typically using 300 \AA of Al. The second oxide is formed as before, by exposure to air during the subsequent processing. Finally, a gold counterelectrode is patterned over the active area by EBL and lift-off as before. The Ge insulation¹¹ separates the base electrode and counterelectrode in all places except for the $20 \times 20\text{-}\mu\text{m}^2$ junction region, thus minimizing leakage and shunt currents.

In order for such a system to display single-electron effects at $T > 4.2\text{ K}$, the dots must have diameters less than 35 nm, assuming hemispherical dots and a temperature threshold of $T = e^2/4kC$.¹² In addition, the dots should be spaced as closely together as possible to limit the parallel shunt current which can flow by directly tunneling from one electrode to the other. This is equivalent to improving the signal-to-noise ratio of the SET effects. The dots must also be very uniform in size and spacing over the entire array. Although it has been predicted¹² that a Coulomb staircase can persist in a system with a standard deviation in particle capacitance as high as 25%, much sharper steps are predicted if the capacitances of the particles are equal.

Therefore, to meet these requirements we have fabricated our arrays using the following parameters. A high-resolution EBL system¹³ was utilized for all exposures. The dots were exposed at 50 kV, in $1400\text{--}2500\text{ \AA}$ of poly(methyl methacrylate) (PMMA) in square arrays with $100\text{--}250\text{-nm}$ spacing, and developed in a 99.7% 1:3 methyl isobutyl ketone (MIBK)/2-propanol, 0.3% methyl ethyl ketone (MEK) solution for 15 s. Typically, the exposure time was limited to about 10 min in order to minimize the effects of current drift and loss of focus.

Figures 2 and 3 show SEM micrographs of typical dot arrays fabricated by the above technique (except that gold

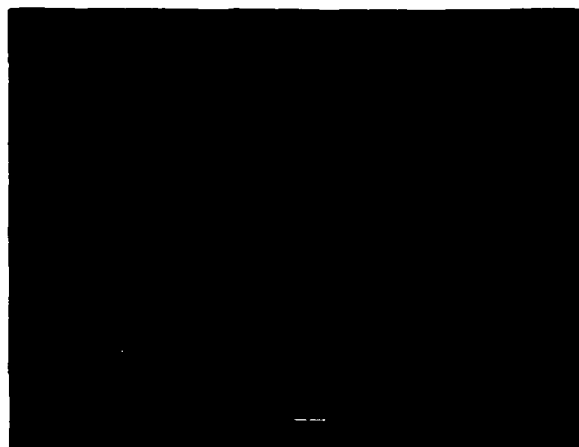


FIG. 2. SEM micrograph of a gold dot array with a 100-nm pitch and 25-30-nm dot diameter.

was used to provide higher contrast in the SEM micrographs). It was found that dot diameters as low as 25 nm could be achieved fairly routinely with extremely successful lift-off results. Even with a dot spacing of 100 nm, nearly 100% lift-off could be achieved. There are several reasons for the high quality of the lift-off. Firstly, the use of a relatively high accelerating voltage (50 kV) reduced proximity effects, thereby creating a more uniform array and forming sharper features in the resist. Next, adding a small amount of MEK to the usual MIBK developer has been shown to enhance the contrast of the developer and improve lift-off characteristics.¹⁴ Finally, the appropriate choice of PMMA layer thickness (up to 2500 \AA) helped to improve the aspect ratio of the resist pattern, making it more suitable for lift-off than thinner resists.

III. LINEAR TUNNEL JUNCTIONS

The other types of structures investigated are single- and multiple-junction systems created by overlapping thin metal lines. Excluding contact pads, these can be created in a single EBL exposure and two successive angled evapora-



FIG. 3. Medium magnification view of a 250-nm-pitch gold dot array showing 100% lift-off in this area.

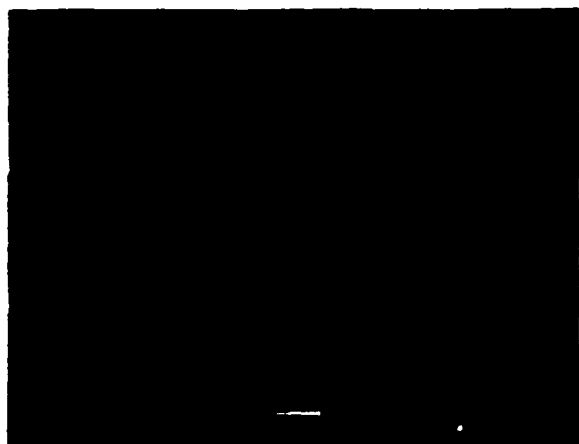


FIG. 4. SEM micrograph of an Al/Al₂O₃/Al tunnel junction with an area of about 30 × 60 nm².

tions with an intermediate oxidation step. The motivation for this particular structure is to create small junctions utilizing a simple and straightforward fabrication technique.

An example of a very small single-tunnel junction is shown in Fig. 4. This pattern simply realizes a junction formed at the intersection of two long, narrow lines. It was created by first exposing a crossed pattern in 2500 Å of PMMA, and developing as discussed above. Next, two angled evaporations of 300 Å of Al were performed with a brief (5-min) oxidation step in air between them. The evaporations were carried out at such an angle (approximately 45°) that only one line would be metallized at a time. As shown in Fig. 4, the two successive angled evaporations resulted in a tunnel junction with an area of 30 × 60 nm².

This same technique was used to create multiple-tunnel junctions in series. An example of a double-tunnel junction is shown in Fig. 5. Here, two parallel lines overlap a small, isolated crossbar which acts as the central electrode. When the resist shadows the incoming evaporant, the metal lines are not exact replicas of the resist features.

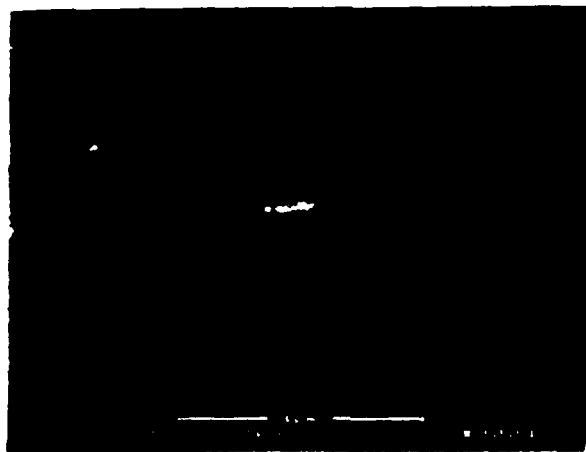


FIG. 5. SEM micrograph of a system containing two linear tunnel junctions in series.

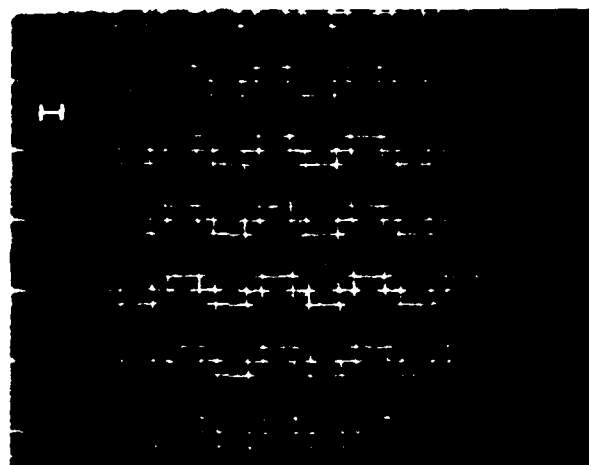


FIG. 6. SEM micrograph of a series of linear tunnel junctions. The bar represents 1 μm.

However, this can be corrected by allowing a certain overlap of the crossing resist lines and shifting the pattern position. It is not sufficient to merely make the actual junctions small, because stray capacitances due to the central bar may still be significant. The most important requirement of this type of structure, therefore, is that the central electrode be as small as possible.

The above processing technique has proven to be very reliable. Figure 6 shows a set of multiple-tunnel junctions demonstrating the feasibility of fabricating large-area structures.

IV. DISCUSSION

The structures which utilize lithographically fabricated dot arrays, as described in Sec. II, merge the precision and high resolution of EBL with the use of isolated particles, thereby offering clear advantages for the study of SET effects. Particle arrays of this type should result in the smallest capacitance attainable in a structure fabricated by conventionally employed lithographic techniques. An advantage of dots used in this system is that no leads are required and the use of dense parallel arrays might also lead to temporal and spatial correlation effects with novel device applications.

To determine the value of capacitance for very small isolated particles fabricated by EBL, the dot can be modeled using a concentric sphere approximation. For a dot of radius r and dielectric thickness d , the capacitance is given by

$$C = 4\pi\epsilon_r\epsilon_0 r [1 + (r/d)]. \quad (2)$$

For the smallest dot size achievable by EBL using PMMA,¹⁵ say, $r = 5$ nm, and typical parameters of $d = 2$ nm and $\epsilon_r = 8$, the corresponding minimum capacitance is $C_{\min} = 1.6 \times 10^{-17}$ F. Conceivably, one might achieve even smaller capacitances by the use of novel lithographic techniques such as those utilizing the scanning tunneling microscope.¹⁶

The properties of the thin tunnel barriers are of major importance to the behavior of SET structures such as those which are based on isolated particle arrays. As stated before, the formation of the tunnel barriers was normally accomplished by utilizing the natural oxidation properties of aluminum to form a thin layer of oxide. However, we found that the oxidation of the base electrode was difficult to control because the sample had to be repeatedly exposed to both air and elevated temperatures (170 °C for PMMA baking) during processing. The result was that all junctions with natural Al_2O_3 barriers were very resistive, thus precluding the observation of SET effects.

An alternative way to control thickness of the tunnel barriers is to form the insulator through deposition rather than oxidation. This can be achieved, for example, by sputter deposition or reactive evaporation of 15–20 Å of Al_2O_3 onto a nonoxidizing metallic substrate. The advantage of these techniques is that the barrier thickness can be precisely controlled and does not change during subsequent processing. Another advantage of using deposited barriers is that the space between the dots is filled with both barrier materials rather than that of the base electrode. This can greatly reduce the parallel shunt current which might flow between the dots,⁸ especially since the resistance of a tunnel barrier increases exponentially with thickness. However, in the event of damage to the oxide during processing, a short circuit between electrodes could result. This is less of a problem with natural barriers which, when exposed to air, tend to "heal" by themselves if damaged.

A final issue for dot arrays fabricated by EBL is the choice of dot metal. For artificial barriers, metals with small grain sizes, such as gold alloys, provide optimum conditions for uniform coating and lift-off. When natural barriers are used, the choice of metal determines the properties of the oxide. If the base electrode barrier is thick, then it is desirable to make the dot barrier thin, and vice versa, so that the resistances of the two barriers are very different. This has been shown to sharpen the steps of the Coulomb staircase in the I - V curve.³ Metals which we have lifted off include Al, Sn, and Pb. Since these metals have

varying oxidation properties, the thickness of the barriers can, in principle, be controlled by appropriate metal selection.

Our approach to the fabrication of linear junctions offers distinct advantages. Firstly, our fabrication technique does not require a multilayer resist system. Secondly, by using perpendicular crossing lines for each junction, no residual patterns are created from the two evaporations, thereby avoiding the presence of unnecessary stray capacitances. Another feature is that the sizes of the junctions do not depend upon the angle of evaporation, but only on the width of the developed lines. In fact, the angle of evaporation does not need to be known precisely because the crossing lines can overlap slightly without greatly increasing the total capacitance of the system. These features make this technique easy to implement while maintaining the capability of producing junctions with dimensions at the limits of electron-beam lithography.

ACKNOWLEDGMENTS

The authors are indebted to S. T. Ruggiero and N. M. Jiarawi for useful conversations. This work was supported in part by the Office of Naval Research and IBM.

- ¹K. K. Likharev, IBM J. Res. Dev. 32, 144 (1988).
- ²T. A. Fulton and G. J. Dolan, Phys. Rev. Lett. 59, 109 (1987).
- ³M. Amman, K. Mullen, and E. Ben-Jacob, J. Appl. Phys. 65, 339 (1988).
- ⁴K. K. Likharev, IEEE Trans. Magn. 23, 1142 (1987).
- ⁵L. J. Geerligs, V. F. Anderregg, P. A. M. Holweg, J. E. Mooij, H. Pothier, D. Esteve, C. Urbina, and M. H. Devoret, Phys. Rev. Lett. 64, 2691 (1990).
- ⁶H. R. Zeller and I. Giaever, Phys. Rev. 181, 789 (1969).
- ⁷L. S. Kuz'min and K. K. Likharev, JETP Lett. 45, 495 (1987).
- ⁸J. B. Barner and S. T. Ruggiero, Phys. Rev. Lett. 59, 807 (1987).
- ⁹G. J. Dolan, Appl. Phys. Lett. 31, 337 (1977).
- ¹⁰G. J. Dolan and T. A. Fulton, IEEE Electron. Dev. Lett. 4, 178 (1983).
- ¹¹E. L. Hu, L. D. Jackel, and R. E. Howard, IEEE Trans. Electron. Devices ED-28, 1382 (1981).
- ¹²K. Mullen, E. Ben-Jacob, R. C. Jaklevic, and Z. Schuss, Phys. Rev. B 37, 98 (1988).
- ¹³G. Bazán, J. Monteiro, and G. H. Bernstein (unpublished).
- ¹⁴G. H. Bernstein and D. A. Hill (unpublished).
- ¹⁵D. Joy, Microelectron. Eng. 1, 103 (1983).
- ¹⁶D. M. Eigler and E. K. Schweizer, Nature 344, 524 (1990).

FIELD EFFECTS IN SELF-CONSISTENT TRANSPORT CALCULATIONS FOR NARROW SPLIT-GATE STRUCTURES*

Henry K. Harbury, Wolfgang Porod, and Craig S. Lent
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556

(Received 19 May 1991)

We study local potential variations due to self-consistent space-charge effects in calculations of coherent transport in narrow split-gate structures. We present a numerical technique based on calculating the Hartree potential from the charge density obtained by solving the two-dimensional effective-mass Schrödinger equation for scattering states, and from the bound charge density obtained from a semi-classical Thomas-Fermi screening model. This method allows us to obtain the local self-consistent potential variations close to scattering centers exposed to an incident flux of electrons.

1. Introduction

Many recent studies have appeared in the literature which present calculations of conductances in narrow split-gate structures that can be realized using advanced fabrication techniques [1]. In these calculations the Schrödinger equation is solved for the channel geometry to obtain transmission coefficients, from which the conductance is obtained. It has been pointed out [2] that self-consistent effects may be significant in these structures, but it has been prohibitively difficult to include them in the numerical calculations.

We present a technique for capturing self-consistent space-charge effects in two-dimensional transport calculations in split-gate structures. We study the self-consistent variations in carrier density and local potentials close to scattering centers exposed to an incident flux of electrons.

2. Numerical Method

Self-consistent electronic states are obtained through an iterative solution of the 2-D Schrödinger equation using current carrying boundary conditions. Any bound electronic charge is approximated by the semi-classical Thomas-Fermi screening model. The Hartree potential due to the electron density is explicitly determined from the total charge density in each iteration, and is used to calculate the electronic states in the subsequent iteration. The self-consistent solution is obtained when the potential dif-

ference between iterations converges to within a desired tolerance.

A. Problem Domain

Our model of the device geometry consists of a discretized calculation domain, Ω_0 , of mesh elements connected to input and output leads, Ω_1 and Ω_2 , as shown in Fig. 1. The current carrying leads attach to the edges of the calculation domain, at $x = 0$ and $x = l$, and extend to infinity. Although the leads are not part of the calculation domain, the charge contained in the leads must be included in the numerical model for the Hartree potential. For simplicity, we assume hard-wall boundary conditions for the conducting channel. A split-gate structure is realized by adding a large fixed potential in the gate regions of the problem. The size of the calculation domain must be sufficiently large that the disturbance due to the constriction has decayed at $x = 0$ and $x = l$. Selecting device geometries that are symmetric across the channel enables us to solve the problem on only one half of the original domain. The resulting bandwidth reduction significantly lowers the computation time.

B. Effective Mass Schrödinger Equation

The 2-D effective-mass Schrödinger equation is repeatedly solved on the discretized spatial domain for the traveling wave "scattering states". V_C denotes the confining

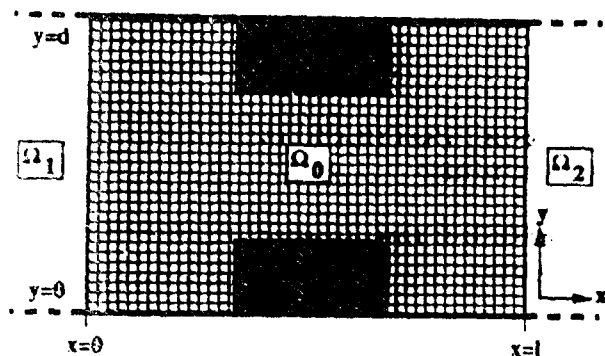


FIGURE 1. Schematic of the channel geometry showing the finite element domain, Ω_0 , and the leads, Ω_1 and Ω_2 . The shaded regions represent the confining gate potentials.

channel potential, V_H represents the self-consistent Hartree potential, and m^* is the electron effective-mass,

$$\left[-\frac{\hbar^2}{2m^*} \nabla^2 + V_c(x, y) + V_H(x, y) \right] \psi_E(x, y) = E \psi_E(x, y). \quad (1)$$

Deep inside the leads, i.e. far from the scattering region, we assume free electron propagation. The electronic wave function can be factorized into a longitudinal plane-wave part, $\exp(ikx)$, and a transverse part with quantized modes, $\chi_n(y)$, which are due to the confining potential of the channel. The scattering eigenstates for injection from the left-hand lead, Ω_1 , in mode m are, therefore, of the form,

$$\psi_E(x, y) = \begin{cases} e^{ik_m x} \chi_m(y) + \sum_n r_{nm} e^{-ik_n x} \chi_n(y) & x \rightarrow -\infty \\ \sum_n t_{nm} e^{ik_n x} \chi_n(y) & x \rightarrow +\infty \end{cases} \quad (2)$$

The index n runs over all possible transverse modes with $E = E_n + (\hbar^2 k_n^2 / 2m^*)$, which includes both traveling and evanescent states.

The finite element method is used to reduce the problem to a large system of coupled linear equations. The Quantum Transmitting Boundary Method [3] is used to implement the finite element boundary conditions appropriate for the current carrying leads.

C. Charge Density

The electron density in the channel due to current carrying states is obtained by explicitly summing over the contributions of all occupied scattering states. We assume the linear response regime and zero temperature. The scattering state electron density is then given by summing over all states with $E \leq E_F$,

$$n_{QM}(x, y) = \sum_{E \leq E_F} |\psi_E(x, y)|^2. \quad (3)$$

The sum includes the wave-functions due to injection from both lead 1 and lead 2. In the linear response regime, injection is symmetric for both leads and this can be used to reduce computation time. We typically find that twenty terms in the summation over electronic states are sufficient.

Bound electronic states can exist inside the scattering region if the total potential dips below its asymptotic value far into the leads. The quantum mechanical electron density obtained for traveling states does not contain any contribution that may arise from bound charge inside the problem domain, Ω_0 . If this contribution is neglected the electron density would be underestimated in the region of a potential well, resulting in a larger attractive Hartree potential and subsequently an unstable runaway condition. As a first order approximation, we choose to model the bound-state electronic charge density with the semi-classical Thomas-Fermi screening model. If the potential in the solution domain is lower than its asymptotic value in the leads, $V_{tot}(x, y) < V_{tot}(0, y)$, then a semi-classical electron density is obtained at the point (x, y) by $n_{TF}(x, y) = \frac{m^*}{\pi \hbar^2} [V_{tot}(0, y) - V_{tot}(x, y)]$, where V_{tot} is the total potential used in solving the Schrödinger equation. This simple approximation over-estimates the total electron density in the region of a potential well because we fill a discrete quantum system with a continuous electron density. The over-estimation, however, will cause the potential well to become more "shallow" in the subsequent iteration resulting in a smaller semi-classical electron density.

To maintain charge neutrality, a compensating positive background charge, ρ_B , is needed. The background charge is obtained by enforcing charge neutrality in the leads. The local disturbance caused by any constriction is assumed to be completely contained within the problem domain, Ω_0 , with negligible effect in the leads. This criterion allows us to model the charge density in the infinitely extending leads as a continuation of the charge density at the edges of the problem domain, $\rho_B = \int_0^d \rho_{el}(0, y) dy = \int_0^d \rho_{el}(l, y) dy$.

This model for the positive background neglects the fact that in most 2DEG systems the positive donors are in a parallel plane isolated from the carriers by a spacer layer. Although we do not model the spacer layer in these calculations, one could, in principle, use an "effective" positive background charge density without changing the algorithm [4].

D. Hartree Potential

The Hartree potential at any location (x, y) is obtained by explicitly summing over all charge contributions in the channel,

$$V_H(x, y) = \frac{1}{4\pi\epsilon} \iint_{\Omega} \frac{\rho(x', y') dx' dy'}{\sqrt{(x-x')^2 + (y-y')^2}} \quad (4)$$

The integral can be broken into three distinct contributions, $V_H = V_H^{\Omega_0} + V_H^{\Omega_1} + V_H^{\Omega_2}$, corresponding to the domain, Ω_0 , and the two leads, Ω_1 and Ω_2 , respectively. The term due to the charge density inside the solution domain, $V_H^{\Omega_0}$, is explicitly calculated by Gaussian quadrature on the finite elements. The components of the Hartree potential due to the infinite leads, $V_H^{\Omega_1}$ and $V_H^{\Omega_2}$, must be, in part, calculated analytically. Using the conditions of charge neutrality across the leads, $\int_0^d \rho(l, y) dy = 0$, and of translational invariance down the leads, $\rho(x > l, y) = \rho(l, y)$, yields

$$V_H^{\Omega_2}(x, y) = -\frac{1}{4\pi\epsilon_0} \int_0^d dy' \rho(l, y') \ln \{ (l-x) + \sqrt{(l-x)^2 + (y'-y)^2} \}. \quad (5)$$

The above integral across the lead is evaluated numerically from the charge density at the edge of the finite element domain, $\rho(l, y)$. We obtain a similar Hartree term due to the charge density in the left lead, $V_H^{\Omega_1}$.

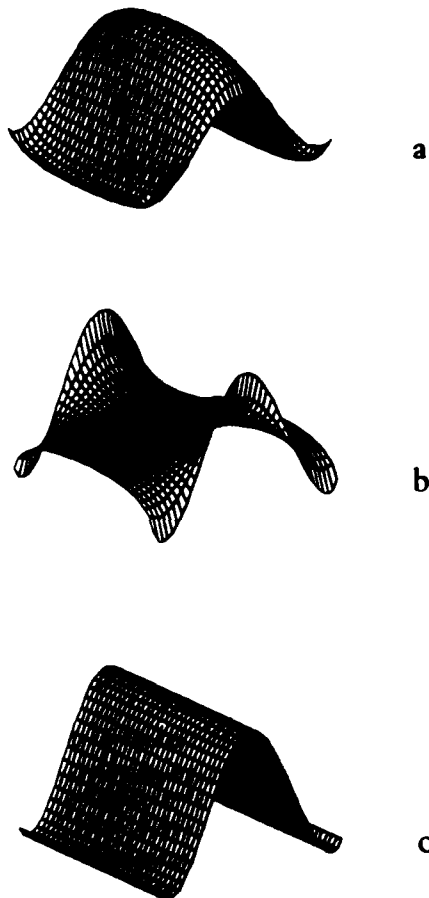


FIGURE 2. The components of the Hartree potential for the open channel after the first iteration: (a) $V_H^{\Omega_0}$, (b) $V_H^{\Omega_1} + V_H^{\Omega_2}$, (c) V_H .

E. Self-Consistency

The resulting Hartree potential is used to modify the total potential on which the Schrödinger equation is solved in the subsequent iteration. The self-consistent potential and charge density are obtained when the difference in the total potential between iterations converges to within a specified tolerance.

3. Results

A. Open Channel

We first investigate the electronic charge density and the Hartree potential for an open, unconfined channel. The solutions must show translational invariance down the channel, namely $\rho(x, y) = \rho(y)$ and $V_H(x, y) = V_H(y)$. This provides a test for the Hartree potential obtained by summing $V_H^{\Omega_0}$, $V_H^{\Omega_1}$, and $V_H^{\Omega_2}$. We choose to study single mode injection by occupying states up to a Fermi energy that is below the second subband in the leads. Fig. 2 shows in part (a) the Hartree potential due to the solution domain $V_H^{\Omega_0}$ and in part (b) the sum of both lead contributions, $V_H^{\Omega_1} + V_H^{\Omega_2}$, after the first iteration. The total Hartree potential, which is the sum of (a) and (b), results in the potential shown in Fig. 2(c) and exhibits the desired trans-

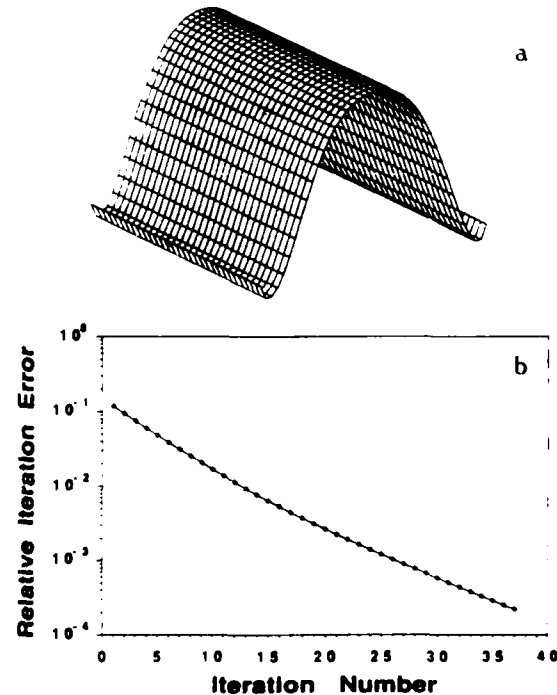


FIGURE 3. Shown is (a) the converged open channel potential, and (b) the iteration error.

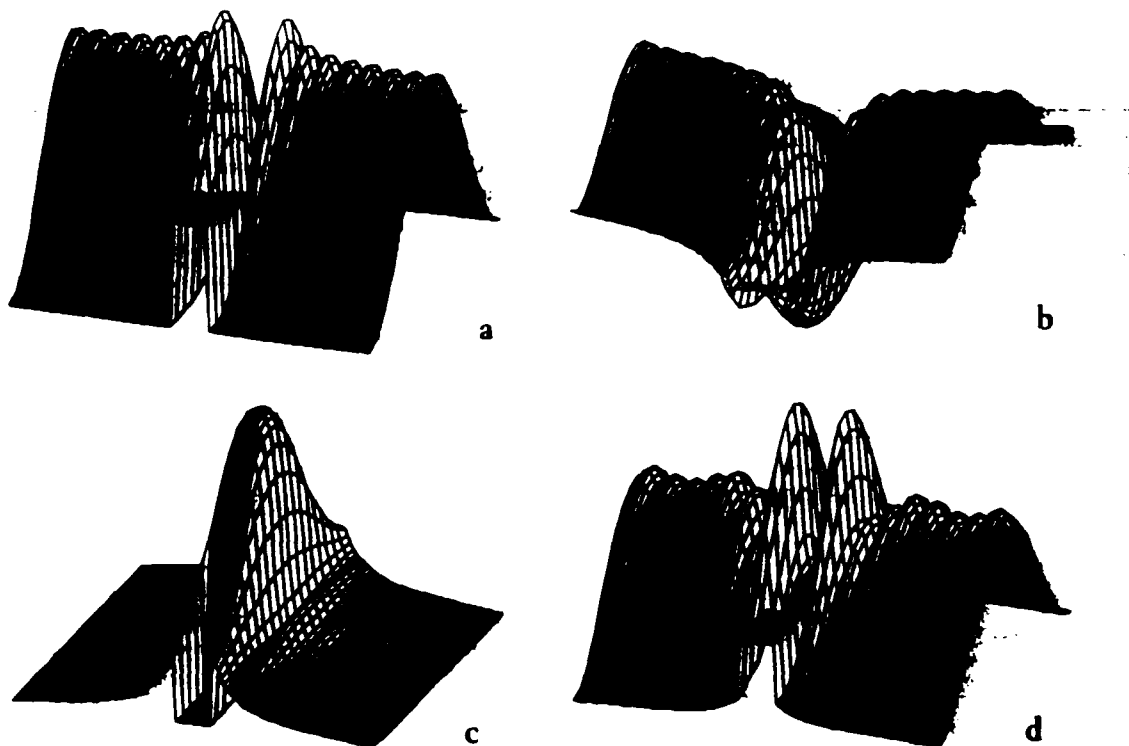


FIGURE 4. Shown is (a) the charge density for the constricted channel (WNW structure) after the first iteration, (b) the new total potential after the first iteration, (c) the semi-classical electron density in the second iteration, and (d) the total charge density after the second iteration.

lational invariance. This clearly shows that the Hartree potential due to charge in the leads exactly compensates for the potential due to charge within the domain, and is a very critical test of our numerical method. Because the total potential in the solution domain, Ω_0 , does not drop below its asymptotic value in the leads, there are no bound electronic states in the open channel. The Hartree potential in Fig. 2(c), obtained from the completely quantum mechanical electron density, is used to modify the total potential with which the Schrödinger equation is solved on the subsequent iteration. Fig. 3(a) shows the converged solution after 37 iterations, and Fig. 3(b) plots the change in the total potential between iterations versus iteration number. The algorithm is stable and convergent for the open channel and, as Fig. 3(a) indicates, the "floor" of the quantum wire is "buckled" by several meV.

B. Constricted Channel

We finally investigate a partially constricted channel, our model for a narrow split-gate structure. We show in

Fig. 4(a) the charge density after the first iteration for single mode injection, and in Fig. 4(b), (c) and (d) the new total potential used in the second iteration, the semi-classical electron density, and the total charge density after the second iteration respectively.

The first iteration charge density far from the constriction asymptotically approaches the case of an undisturbed system. The leakage of charge between the split gates in the first iteration is small enough that the charge density in Fig. 4(a) resembles the solution of the totally constricted channel [5]. The smaller electron density in the gate region in Fig. 4(a) results in the formation of a potential well through the Hartree term used to obtain the total potential shown in Fig. 4(b). This potential well in the gate region will, in subsequent iterations, cause an increase in the local electron density near the scattering center. This results both from an increase in occupied traveling states through the gate region and from the semi-classical model for the bound electron density, shown in Fig. 4(c), since the potential in the gate region dips below the po-

tential far into the leads. The total charge density obtained in the second iteration, due to both the electron density of the traveling states and the semiclassical bound electron density, is shown in Fig. 4(d). The increase in the electron density near the scattering center is clearly visible. Although not fully converged, it is clear from these calculations that the oscillations in the local field near the scattering center due to self-consistent space charge effects could be significant in transport calculations since they will effect the transmission amplitudes of the structure. The oscillations in the potential due to a scattering center in the 700Å channel can extend up to a half micron into the leads, and are on the order of a few meV. Although the potential well in the scattering region does not become deeper in later iterations, we have found it difficult to achieve convergence for split-gate structures. We are currently investigating various convergence schemes and more work is needed on this problem.

4. Conclusion

We have presented a method which incorporates self-consistent space charge effects in transport calculations for narrow split-gate structures. Our method uses an iterative solution of the effective-mass Schrödinger equation, with a semi-classical model for bound electronic charge, and explicit calculation of the Hartree potential. This method

allows us to investigate the local self-consistent potential variations close to scattering centers exposed to an incident flux of electrons.

For the case of an open channel, or quantum wire, we have demonstrated the validity of our method and achieved a fully convergent solution. We presented some results for a partially constricted channel, or narrow split-gate structure, but for this case convergence has been difficult to achieve. More work is needed to ensure convergence of the self-consistent cycle for general split-gate structures.

5. References

- [1] *Nanostructures Physics and Fabrication*, eds. M. A. Reed and W. P. Kirk, (Academic Press, 1989).
- [2] L. Escapa and N. Garcia, *Appl. Phys. Lett.* **56**, 901 (1990).
- [3] C. S. Lent and D. J. Kirkner, *J. Appl. Phys.* **67**, 6353 (1990).
- [4] J. A. Nixon and J. H. Davies, *Phys. Rev. B* **41**, 7929 (1990).
- [5] H. K. Harbury, W. Porod, C. S. Lent, *Proceedings of the Seventh International Conference on the Numerical Analysis of Semiconductor Devices and Integrated Circuits, NASECODE-7*, p. 136, ed. by J. J. H. Miller, (Front Range Press, 1991).

ON THE ATTAINMENT OF OPTIMUM DEVELOPER PARAMETERS FOR PMMA RESIST

Gary H. Bernstein⁽¹⁾ and Davide A. Hill⁽²⁾

⁽¹⁾Department of Electrical Engineering

⁽²⁾Department of Chemical Engineering

University of Notre Dame

Notre Dame, IN 46556

(Received 19 May 1991)

Because of the uniquely high resolution offered by poly(methyl methacrylate) (PMMA) for practically all forms of lithography, its technological significance remains strong. We have found significantly higher contrast developers for PMMA which have shown improved resolution, less susceptibility to proximity effects and better lift-off properties with no loss of sensitivity to exposure. We report data which demonstrate an improvement in contrast by more than 35% with the addition of as little as 1.3% methyl ethyl ketone (mek) to other common developers. We discuss a model which explains the observed contrast enhancement and predicts that for small changes, an increase in either mek concentration, temperature or develop time increases contrast. We include experimental evidence in support of temperature and mek related improvements.

1. Introduction

Of all materials utilized in the fabrication of nanostructures, poly (methyl methacrylate) (PMMA) is among the most important. PMMA remains so popular as a positive resist that virtually all laterally defined quantum devices rely on PMMA for patterns below 100 nm regardless of the lithographic technique employed. In fact, PMMA is the only proven positive resist usable into the 10 nm regime. (Contamination resist is utilized for etch masks in this dimensional regime, but is slower and more difficult to work with.)

The typical manner in which PMMA is employed is that of exposure to electrons, X-rays or deep UV light followed by development, metallization and lift-off. The resolution of the resulting metal pattern depends on both the width of the defined line and the quality of the edge profile in the resist. In turn, edge profile depends on the contrast, γ , of the resist/developer system. Contrast is defined as the maximum slope of the curve of normalized resist thickness remaining after development as a function of the log of the exposure dose. As contrast increases,

such properties as lift-off reliability, proximity effect immunity and, to some extent, resolution, are all improved.

We have found that several parameters can increase the contrast of PMMA. First, we found that adding a very small percentage of methyl ethyl ketone (mek) to other common developers effects a drastic improvement in contrast, dimensional consistency and lift-off quality. Second, we have found that, over a narrow temperature range, an increase in temperature also improves contrast. Finally, as discussed in further sections, we propose that an increase in develop time is also predicted to improve contrast.

2. Experimental

Contrast results for a variety of developer formulations were determined using a standard technique^(1,2). Using electron beam lithography (EBL) at 40 kV, we exposed 25 50 μm x 50 μm squares in 950,000 amu PMMA, 0.4 to 1 μm thick with a series of doses, developed for various times and temperatures, and determined the normalized resist

thicknesses by surface profilometry using a Sloan Dektak II profilometer. The quality of the data was excellent, and an example of the raw data is given in Fig. 1. Variations in resist thickness were shown to have no effect on the results. Experimental parameters and contrasts are summarized in Table I.

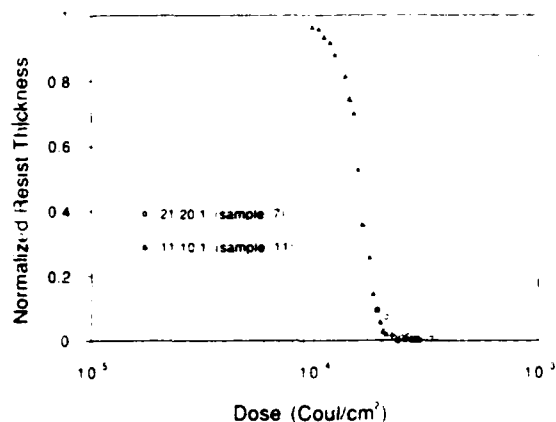


Figure 1: Example of data from which contrast is calculated.

The primary ingredients of the developer mixtures were chosen from solutions commonly employed as developers for PMMA. These are methyl ethyl ketone ("mek"):ethyl alcohol (EtOH) 26.5:73.5 (mixture referred to here as "MEK"), methyl isobutyl ketone (mibk):isopropyl alcohol (IPA) 1:3 (mixture referred to here as "MIBK"), and 2-ethoxy ethanol (Cellosolve or "cs"):methanol (MeOH) (mixture referred to here as "CS"). These mixtures were chosen in their complete forms due to their known properties as developers of PMMA. It should be noted that solubility properties of mixtures of several solvents are extremely difficult or impossible to predict, and such an exercise will not be attempted here. In subsequent sections we offer a qualitative explanation of the role of the harshest component, mek, in improving contrast and offer guidelines toward future improvements in contrast.

3. Results and Discussion

For mixtures given in Table I, we measured contrast and plotted the results as a function of inverse temperature as shown in Fig. 2. A trend of increasing contrast with temperature in the range

Table I: Summary of contrast experiment data.

Sample I.D.	volume percent MIBK	volume percent CS	volume percent MEK	Develop time (sec)	Temp °C	γ
1	0	100	0	5	19	5.7
2	0	100	0	5	21	7.0
3	100	0	0	45	19	7.0
4	100	0	0	45	21	9.3
5	66.4	32.0	1.6	10		7.7
6	66.4	32.0	1.6	10		7.0
7	49.6	48.0	2.4	10		8.2
8	49.6	48.0	2.4	10		8.5
9	50.1	45.4	4.5	10	18	7.9
10	50.1	45.4	4.5	10	18	7.9
11	50.1	45.4	4.5	10	19	9.4
12	50.1	45.4	4.5	10	19	9.4
13	50.1	45.4	4.5	10	19	9.1
14	50.1	45.4	4.5	10	19	8.8
15	0	95	5	5	21	12.8
16	0	0	100	2	19	4.7

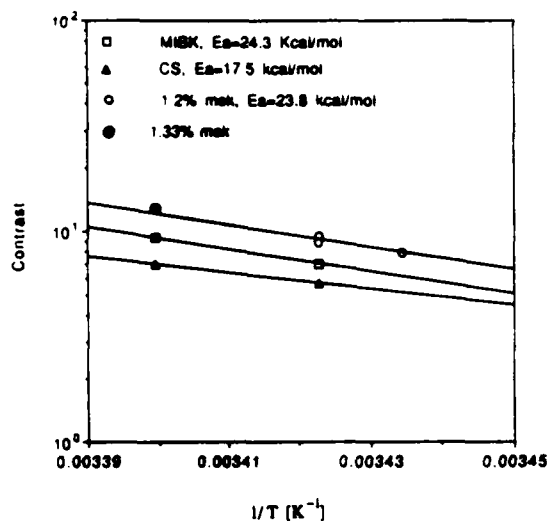


Figure 2: Arrhenius plot of contrast for various formulations.

between 18°C and 21°C is clearly evident. This range is especially important since room temperature variations coupled with evaporative cooling of the developer solvents are shown to be especially critical. It can be shown²⁰ that γ can be expressed as

$$\gamma = \left[r_d \cdot \frac{\partial \log r_d}{\partial \log M} \cdot \frac{d \log M}{d \log D} \cdot \frac{\Delta t_d}{h_0} \right]_{\max} \quad (1)$$

where r_d is the average rate of dissolution, M is the molecular weight, D is the dose, Δt_d is the development time and h_0 is the initial resist thickness. The first term, r_d , depends on $\exp(-E_a/RT)$ where E_a is the activation energy, R is the gas constant and T is the temperature. This theory also predicts that γ depends linearly on develop time. This has not yet been investigated and remains for future work on this topic.

Figure 3 gives the dependence of contrast on percent mek (pure methyl ethyl ketone component) normalized to 19°C. The CS data are omitted since the activation energy of Cellosolve is significantly different from that of MIBK and MIBK/MEK mixtures (see Fig. 1) and therefore do not normalize to the plot. Note that data corresponding to 21°C for pure MIBK and that of 18°C at 1.19% mek have collapsed onto the straight line and their multiplicity is obscured.

An explanation of the time, temperature and mek concentration dependence follows. Figure 4 is the qualitative dependence of dissolution rate on the

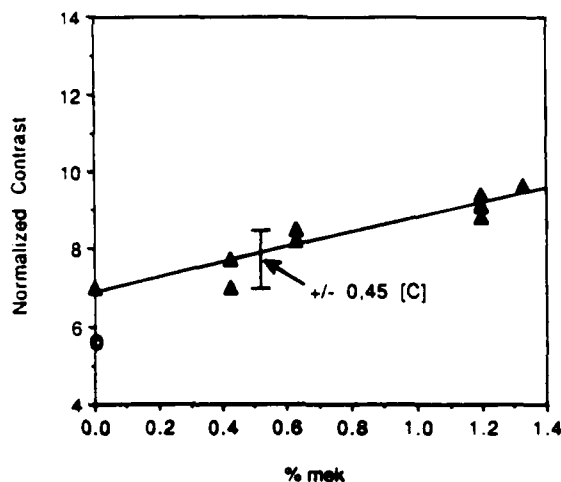


Figure 3: Plot of contrast as a function of % mek formulation normalized to 19°C.

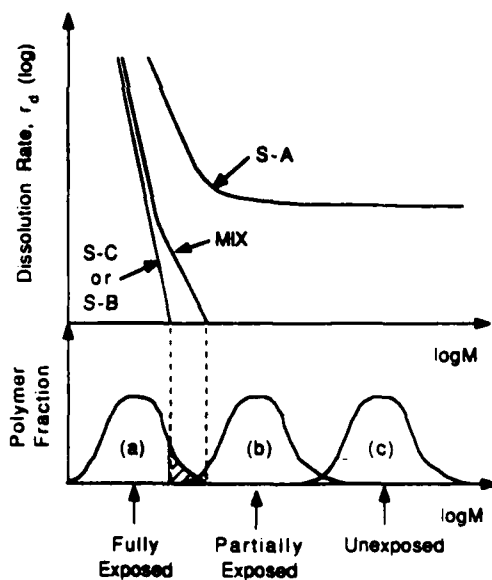


Figure 4: Dependence of dissolution rate on temperature, time and solvent strength superimposed on fragmented molecular weight distribution.

strength of a solvent, time or temperature as a function of molecular weight, M . This dependence has been demonstrated by Greeneich for MIBK/IPA solutions⁽⁴⁾. Below this curve is the dependence of the M distribution after three exposure doses⁽²⁾. First, as any of the above parameters is slightly increased, the solvent quality changes in such a manner as to dissolve more efficiently the higher M tails remaining

after normal exposures⁽¹⁾. For the case of very slight amounts of MEK added to MIBK, the mek acts independently of the MIBK to remove efficiently high M components left behind by the MIBK, thus improving contrast, side wall profiles and proximity effect immunity without decreasing sensitivity. The effect on lower M components is not increased enough to affect deleteriously the contrast⁽²⁾.

The above results suggest a way in which developers might be further formulated to increase contrast even higher. We suggest that solvents be chosen for their range of effectiveness in dissolving various molecular weights of PMMA. Assuming these solvents retain their basic properties in solution, the solvents could be mixed in the percentages which reflect the percentages of molecular weights of the PMMA remaining after exposure (as shown in Fig. 4). These tailored developers would be strong enough to remove only the proportions of the PMMA in the particular ranges of M affected by the individual components but would not act to remove higher M components in adjoining, less exposed areas or in unexposed areas. In a very simplified way, this is what has been accomplished by the addition of mek, a much stronger solvent for PMMA, to mibk. Clearly, much work in this area is still to be done.

4. Summary

We have demonstrated that new developer

solutions containing small amounts of a strong developer mixed with a weaker one can enhance all aspects of PMMA development with no deleterious side effects. In addition, we demonstrated experimentally that good control of temperature is critical for consistent contrast control, where in fact, at least slightly higher temperatures are desirable. New directions for research in this area were suggested.

Acknowledgements

The authors wish to thank Wen-Ping Liu for sample preparation and D. K. Ferry for helpful discussions. This work was supported in part by IBM and ONR (G.H.B.) and the University of Notre Dame.

References

1. M. A. Mohsin and J. M. G. Cowie, *Polymer* **29**, 2130 (1988).
2. V. K. Sharma, R. A. Pethrick and S. Affrossman, *Polymer* **23**, 1732 (1982).
3. D. A. Hill and G. H. Bernstein, unpublished.
4. J. S. Greeneich, *J. Electrochem. Soc.: Sol. St. Sci. and Techn.* **122**, 970 (1975).
5. S. Mackie and S. P. Beaumont, *Sol. St. Techn.*, August, 117 (1985).

Numerical calculation of hybrid magneto-electric states in an electron waveguide

S. Chaudhuri and S. Bandyopadhyay

Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

(Received 30 October 1991; accepted for publication 6 December 1991)

We have performed a numerical calculation of the energy dispersion relation of hybrid magneto-electric states (both propagating and evanescent) in an electron waveguide subjected to a magnetic field. Our results are considerably different from those obtained through the Bohr-Sommerfeld quantization condition. We have also calculated the density of the magneto-electric states as a function of energy and the velocity versus energy relationships. Finally, we show how the wavefunctions of these states evolve with increasing magnetic field from particle in a box states to edge states. These results are useful in the analysis of numerous recent magnetotransport experiments performed in electron waveguides.

Electron waveguides (quasi-one-dimensional wires in which carrier transport is partially ballistic) has been the focus of many recent theoretical and experimental investigations.¹ Of particular interest in these structures is the nature of magneto-electric states that are formed under an applied transverse magnetic field.² These states are important in understanding numerous magnetotransport experiments such as the conductance oscillations of quasi-one-dimensional electron gases in a magnetic field,³ the integer quantum Hall effect,⁴ etc. In the past, magneto-electric subbands and their wave functions were calculated in an approximate manner using analytical functions such as Weber or hypergeometric functions.⁵ In this paper, we have calculated the wave function, the energy dispersion relation, the velocity versus energy relationships, and the density of these states starting from the Schrödinger equation. Our calculation is described below.

We consider an electron waveguide (a quasi-one-dimensional structure) as shown in Fig. 1. The magnetic field is directed along the z direction along which the confinement is complete in that only one transverse subband is occupied. The Schrödinger equation describing electrons in this structure is

$$\frac{(\mathbf{p} - e\mathbf{A})^2}{2m^*} \psi(x, y) + V(y)\psi(x, y) = E\psi(x, y), \quad (1)$$

where \mathbf{A} is the magnetic vector potential, m^* is the electron's effective mass, and $V(y)$ is the electric confining potential in the y direction. The simultaneous presence of a magnetic potential and an electric potential hybridizes the electronic states causing them to be magneto-electric states.

The potential $V(y)$ is chosen to be consistent with hardwall boundary conditions

$$\begin{aligned} V(y) &= 0 \quad |y| < d \\ &= \infty \quad |y| > d, \end{aligned} \quad (2)$$

where $2d$ is the width of the structure in the y direction. We choose the Landau gauge

$$\mathbf{A} = (-By, 0, 0), \quad (3)$$

where B is the z directed magnetic flux density.

The wave function ψ can be written as

$$\psi(x, y) = e^{ik_x x} \phi(y), \quad (4)$$

where k_x is the x directed wave vector and $\phi(y)$ satisfies the eigenequation

$$\begin{aligned} \frac{\partial^2 \phi(y)}{\partial y^2} + \frac{2m^*}{\hbar^2} E \phi(y) - \left(\frac{y}{l}\right)^2 \phi(y) + 2\frac{y}{l} k_x \phi(y) \\ - k_x^2 \phi(y) = 0 \end{aligned} \quad (5)$$

with l being the magnetic length given by $l = \sqrt{\hbar/eB}$. To find the wave function $\phi(y)$ of the magneto-electric states, we have to solve the above equation subject to the boundary conditions

$$\phi(y = d) = \phi(y = -d) = 0 \quad (6)$$

which follow from Eq. (2).

In order to find the energy dispersion relation E vs k_x for the magneto-electric states, we have to find the values of the wave vector k_x for a given energy E satisfying Eqs. (5) and (6). We then repeat this for various values of E to obtain the energy dispersion relation. Unfortunately, this is not straightforward since Eq. (5) is not an eigenequation in k_x for a given E due to the fact that it is nonlinear in k_x .

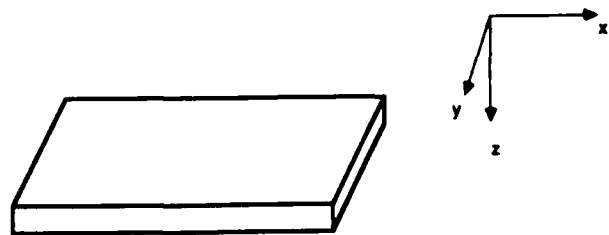


FIG. 1. An electron waveguide subjected to a magnetic field along the z axis.

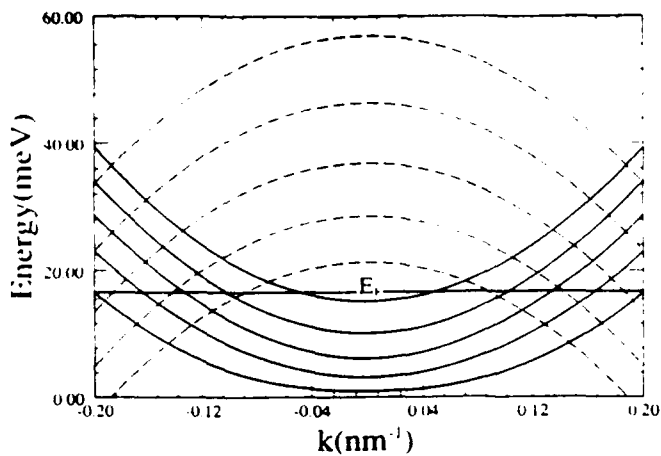


FIG. 2. Energy dispersion E vs k_x of magneto-electric subbands in a waveguide subjected to a magnetic field. The waveguide transverse width is 100 \AA and the magnetic field is 1 T . The Fermi energy is taken to be 17 meV . If the subband bottoms are below the Fermi energy, the states are propagating states and have real wave vectors; otherwise they are evanescent states and have imaginary or complex wavevectors. For the parameters (well width and magnetic field) chosen, the lowest evanescent states have purely imaginary wave vectors. The evanescent states are shown with negative curvatures. The energy separation between the subband bottoms depends on the magnetic field as well as the width of the waveguide.

We therefore have to convert Eq. (5) into an eigenequation in k_x using the following transformation.

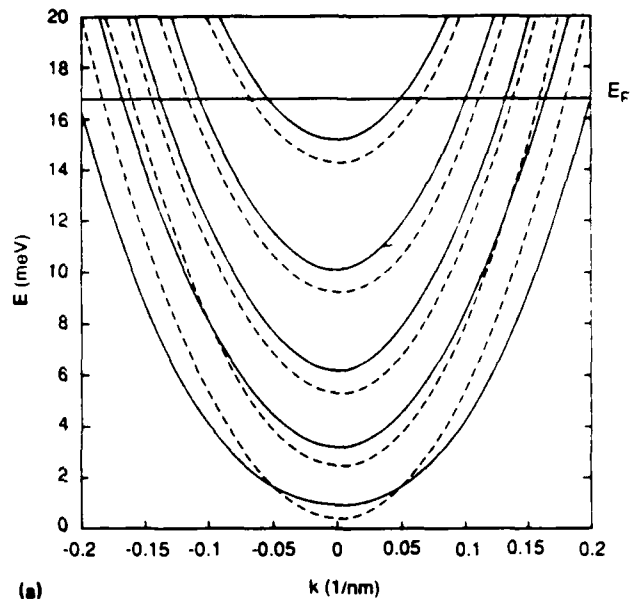
Let

$$\zeta(y) = k_x \phi(y) \quad (7)$$

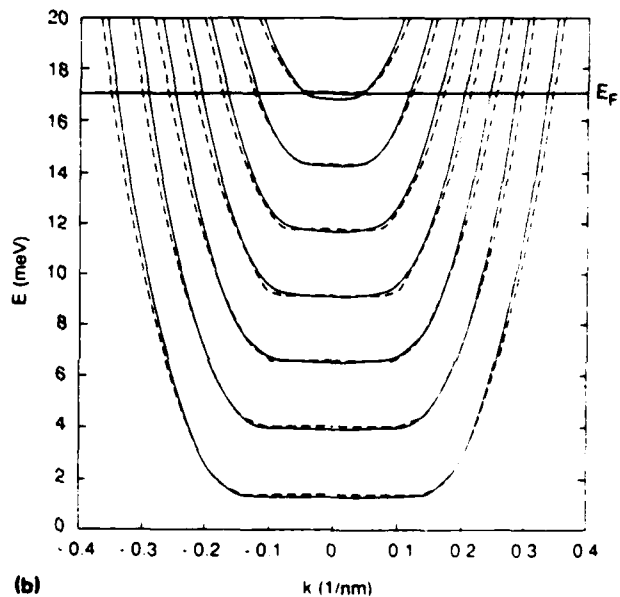
Equation (5) can now be recast as

$$\left(\frac{\partial^2}{\partial y^2} + \frac{2m^*}{\hbar^2} E - \left(\frac{y}{\bar{p}} \right)^2 \right) \begin{pmatrix} \phi \\ \zeta \end{pmatrix} = k_x \begin{pmatrix} \phi \\ \zeta \end{pmatrix}. \quad (8)$$

We discretize Eq. (8) within a finite difference scheme breaking the y domain into N grid points where N is some suitably large integer. We can now readily solve for N different k_x eigenvalues ($k_x^1, k_x^2, \dots, k_x^N$) for any given energy E and find the corresponding eigenvectors which give the wave function $\phi_n(y)$. Each value of n corresponds to a magneto-electric subband. This gives us the energy E versus the k_x^2 ($n = 1, 2, 3, \dots, N$) relation, or the dispersion relation for N subbands. We vary E up to an arbitrary maximum which we call the "Fermi energy" E_F . If the bottom of a subband is above E_F , then the corresponding state is an evanescent state; otherwise, it is a propagating state. Evanescent states have imaginary or complex wave vectors (k_x values) whereas propagating states have real k_x values. It is interesting to note that while in the absence of a magnetic field evanescent states can only have imaginary wave vectors, in the presence of a magnetic field they can have both imaginary and complex wave vectors. The states with complex wave vectors can interfere since their wave functions are complex and they have nonzero current densities associated with them *even though they are evanescent*. It is interesting to note that because of this feature one could



(a)



(b)

FIG. 3. Energy dispersion relations (for the propagating states only) obtained from our results (solid curves) and the Bohr-Sommerfeld quantization rule (broken curves). (a) The waveguide width is 1000 \AA and the magnetic field is 1 T ; (b) the waveguide width is 2000 \AA and the magnetic field is 1.5 T .

observe, say, the magnetostatic Aharonov-Bohm effect with the evanescent states but not the electrostatic Aharonov-Bohm effect.

In Fig. 2 we show the energy dispersion relations including both propagating and evanescent states. In Figs. 3(a) and 3(b) we compare the dispersion relations of the propagating states with those obtained from the Bohr-Sommerfeld quantization rule.^{6,7} We find that there is a significant difference between the two results. In Fig. 4, we show the velocity versus energy relations for different subbands. The velocity in the n th subband is defined as $v_n = (1/\hbar)(\partial E_n / \partial k_x)$.⁸ In Fig. 5, we show the density of states versus energy where the density of states is defined as $\Sigma_n(4/\hbar v_n)\theta(E_n - E_n^0)$ with θ being the Heaviside step

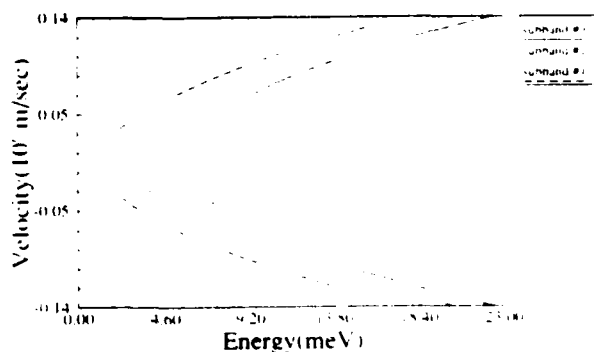


FIG. 4. Velocity vs energy relationships for the magneto-electric states in a waveguide width of 1000 Å and a magnetic field of 1 T.

function and E_n^0 the energy at the bottom of the n th subband. The summation is carried out over the propagating states only. The density of magnetoelectric states is useful in calculating the optical absorption spectra of a quantum wire subjected to a magnetic field, or the scattering rates for hot electrons in the presence of a magnetic field, and various other quantities.

Finally, in Figs. 6(a), 6(b), and 6(c), we show how the probability density $|\phi(y)|^2$ of the lowest subband at the energy $E = E_F$ evolves with increasing magnetic field. At zero magnetic field, the state is a particle in a box state, whereas at high magnetic fields, it peaks near the edges of the waveguide and becomes an edge state. For a given sign of the wave vector k_x , the probability density peaks near one of the edges of the waveguide whereas for the opposite sign of k_x , it will peak near the opposite edge. This behavior can be understood as skewing of the wavefunction towards one of the edges due to the Lorentz force applied on the electrons.

In conclusion, we have shown in this paper the nature of magneto-electric states in an electron waveguide. These states have been invoked to explain several experiments and phenomena such as magnetoconductance oscillations in electron waveguides and the quantum Hall effect. We

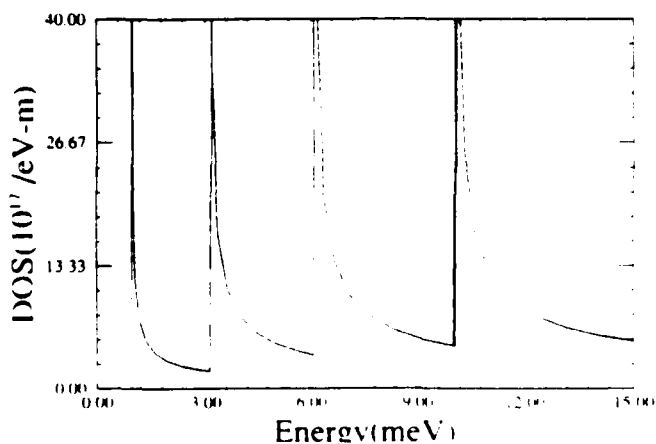


FIG. 5. Density of magneto-electric states in a 1000 Å wide waveguide at a field of 1 T.

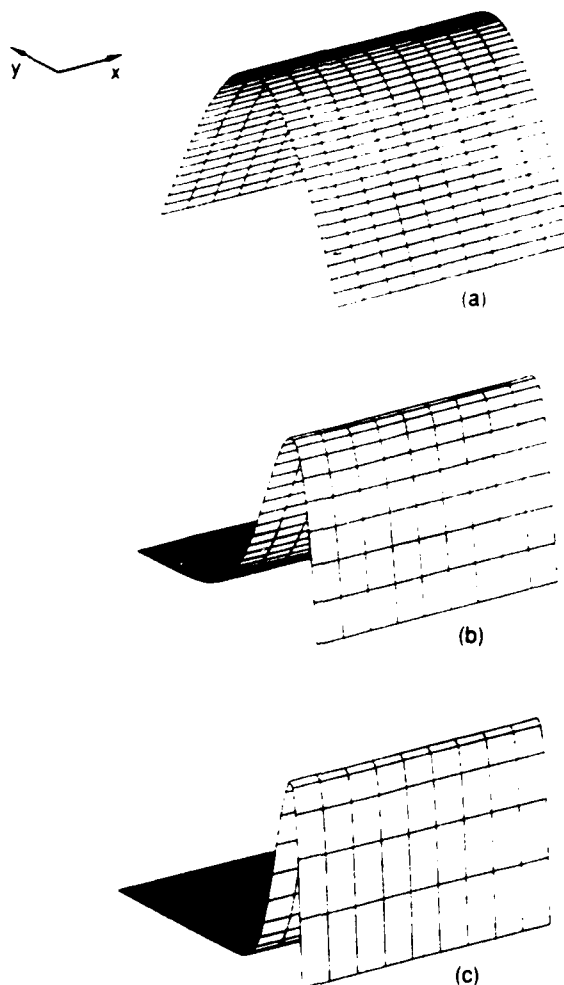


FIG. 6. Probability density in the lowest magneto-electric subband at the Fermi energy of 17 meV. The waveguide width is 1000 Å. (a) The magnetic field is zero; (b) the magnetic field is 1 T; (c) the magnetic field is 10 T.

believe that our results will be useful in the analysis of these and other experiments dealing with magneto-electric states.

This work was supported by the U.S. Air Force Office of Scientific Research under grant No. AFOSR 91-0211 and by the U.S. Office of Naval Research under grant number N00014-91-J-1505.

¹ For recent reviews, see, for example, *Nanostructure Physics and Fabrication*, edited by M. A. Reed and W. P. Kirk (Academic, Boston, 1989).

² See, for example, W. Zawadzki, S. Klahn, and U. Merkt, *Phys. Rev. B* **33**, 6916 (1986); J. Yoshino, H. Sakaki, and T. Hotta, *Surf. Sci.* **142**, 326 (1984).

³ S. B. Kaplan and A. C. Warren, *Phys. Rev. B* **34**, 1346 (1986).

⁴ M. Büttiker, *Phys. Rev. B* **38**, 9375 (1988).

⁵ A. H. McDonald and P. Streda, *Phys. Rev. B* **29**, 1616 (1984).

⁶ C. W. J. Beenakker, H. van Houten, and B. J. van Wees, *Superlatt. and Microstructures* **5**, 127 (1989).

⁷ C. W. J. Beenakker and H. van Houten, *Solid State Phys.* **44**, 228 (1991).

Transmission resonances and zeros in quantum waveguides with resonantly coupled cavities

Wolfgang Porod, Zhi-an Shao, and Craig S. Lent

Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

(Received 19 May 1992; accepted for publication 24 June 1992)

We report on the existence of transmission zeros in quantum waveguide structures with resonantly coupled cavities. Such zeros do *not* occur in the usual double-barrier resonant tunneling systems. For quantum waveguides, the transmission probability exhibits pairs of poles and zeros in the complex-energy plane. The observed sharp structure of the transmission resonances and zeros can be understood in terms of these zero-pole pairs.

Electronic transport in ultrasmall semiconductor structures resembles wave propagation in waveguides,^{1,2} and device applications based on this analogy to microwave devices have been proposed. The transmission amplitude in these systems exhibits a rich structure which is related to resonance phenomena.³⁻⁵ In this letter, we demonstrate how quasibound states in resonantly coupled cavities give rise to zero-pole pairs in the complex-energy plane. Transmission zeros are unique to quantum waveguide structures, and are absent for double-barrier resonant tunneling.

Resonances have long been studied in transmission through double-barrier resonant-tunneling structures.^{6,7} It is well known that these resonances are related to the existence of quasibound states in the quantum-well region. Within the Breit-Wigner formalism, a quasibound state at energy E_0 and decay time $\tau = \hbar/\Gamma$ will give rise to a transmission resonance with a Lorentzian line shape, $T(E) = \frac{1}{4}\Gamma^2/[(E-E_0)^2 + \frac{1}{4}\Gamma^2]$. In the complex-energy plane, this corresponds to a simple pole of the transmission amplitude at the complex energy $z = E_0 - i\Gamma/2$.⁸

It is instructive to demonstrate why a quasibound state gives rise to a pole in the complex-energy plane. In analogy to an optical Fabry-Perot resonator, the total transmission amplitude across both barriers, \mathcal{T}_{RL} (from left to right), may be related to partial transmission and reflection coefficients at each barrier:⁹

$$\mathcal{T}_{RL} = t_{RW}(e^{ikL} + e^{ikL}r_R e^{ikL}r_L e^{ikL} + e^{ikL}r_R e^{ikL}r_L e^{ikL}r_R e^{ikL}r_L e^{ikL} + \dots)t_{WL} \quad (1)$$

$$= t_{RW} \frac{1}{e^{-ikL} - r_R e^{ikL}r_L} t_{WL}. \quad (2)$$

Here, t_{WL} denotes the transmission amplitude from the left to the well region, and r_R is the reflection coefficient at the right barrier. The phase accumulated by an electron with energy E moving from one side of the quantum-well resonator of length L to the other is represented by the phase factor $\exp(ikL)$, where $k = \sqrt{(2m^*E)/\hbar^2}$. Poles occur at those complex energies for which the denominator vanishes. For real-value reflection amplitudes, it is an easy matter to see that this occurs for wave numbers k whose real part is an integer multiple of π/L , i.e., for quasibound states. Note that this geometric series does not possess zeros.

As an illustration of the above arguments, we show in Fig. 1 the transmission amplitude in the complex-energy plane for a double-barrier resonant-tunneling structure which is schematically depicted in the inset. This structure supports true bound states for $E < 0$, double-barrier resonances for $0 < E < V_0$, and continuum resonances for $E > V_0$. Figure 1(a) shows the transmission coefficient on the real-energy axis ($E > 0$), while Fig. 1(b) depicts a contour plot of the absolute value of the transmission amplitude in the complex-energy plane. Note that transmission resonances and poles occur at essentially the same energies.

We now proceed to the study of transmission in quantum waveguides in the presence of coupled resonators. The prototypical resonator structure we adopt is the resonantly coupled stub, as schematically shown in Fig. 2. The resonant cavity is created by closing off the side arm in a three-way branch [Fig. 2(b)]. The new feature here, with respect to quasi-one-dimensional double-barrier resonant-tunneling is the existence of wire branches which inherently makes this problem spatially two-dimensional.

We start out by considering transmission through the three-way splitter [Fig. 2(a)]. This problem has been studied by several workers,¹⁰⁻¹² and we follow them in choosing a scattering matrix approach which connects the out-going and in-coming waves in each wire branch.

$$\begin{pmatrix} O_L \\ O_R \\ O_S \end{pmatrix} = \begin{pmatrix} r_L & t_{LR} & t_{LS} \\ t_{RL} & r_R & t_{RS} \\ t_{SL} & t_{SR} & r_S \end{pmatrix} \begin{pmatrix} I_L \\ I_R \\ I_S \end{pmatrix}. \quad (3)$$

The elements of the scattering matrix are constrained by unitarity. The reflection and transmission amplitudes are the elements of a Shapiro-Büttiker matrix,^{10,13} and their specific values depend upon the detailed geometry of the waveguide and junction. Typically, the elements of the scattering matrix are slowly varying functions of energy.

Making the sidearm into a resonator forces a standing wave in the stub [Fig. 2(b)]. This imposes an additional relationship between the out-going and in-coming amplitudes in the side branch, O_S and I_S , respectively,

$$O_S = \lambda(E)I_S, \quad (4)$$

where $\lambda = e^{i\Phi(E)}$ and the phase angle $\Phi(E)$ is a property of the resonator.⁵

The transmission and reflection amplitudes for the quantum waveguide with a resonantly coupled stub, \mathcal{T}

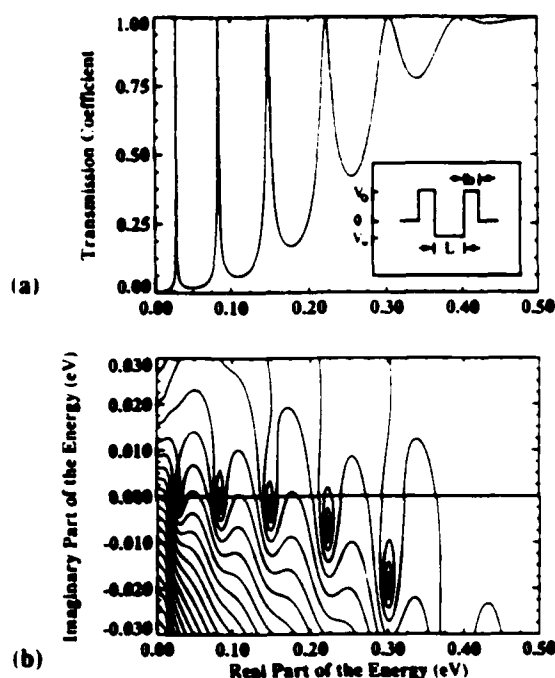


FIG. 1. Transmission amplitude for a double-barrier resonant-tunneling structure, which is schematically depicted in the inset ($V_0=0.2$ eV, $V_w=0.1$ eV, $L=30$ nm, and $b=5$ nm); (a) shows the transmission coefficient on the real-energy axis, and (b) shows a contour plot of the absolute value of the transmission amplitude in the complex-energy plane.

and \mathcal{R} , may now be determined by combining Eqs. (3) and (4). For the case of transmission from left to right, they are given by

$$\mathcal{T}_{RL} = t_{RL} + \frac{t_{RS}t_{SL}}{\lambda - r_S}, \quad \text{and} \quad \mathcal{R}_L = r_L + \frac{t_{LS}t_{SL}}{\lambda - r_S}. \quad (5)$$

Note that the second terms contain a resonance denomi-

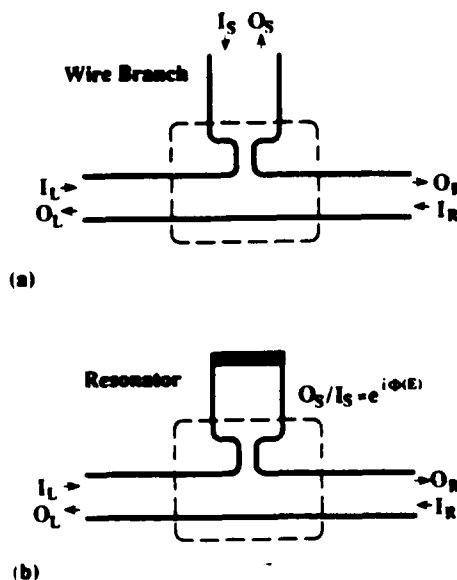


FIG. 2. Schematic drawing of a waveguide with a resonantly coupled cavity. (a) shows a wire branch with incoming and outgoing waves outside the junction region, which is indicated by the dashed box; (b) shows a resonant stub which is obtained by closing off the sidearm.

nator, which is reminiscent of double-barrier resonances, Eq. (2). An expansion of (5) yields

$$\mathcal{T}_{RL} = t_{RL} + t_{RS} \left(\frac{1}{\lambda} + \frac{1}{\lambda} r_S \frac{1}{\lambda} + \frac{1}{\lambda} r_S^2 \frac{1}{\lambda} + \dots \right) t_{SL}, \quad (6)$$

which allows the following interpretation of the transmission and reflection amplitudes for the resonantly coupled stub. Each multiple reflection in the stub contributes a factor r_S for reflection back into the sidearm and a phase factor $1/\lambda$ for the roundtrip (up and down). The sum of all multiple reflections results in a geometric series, very much like a Fabry-Perot resonator. Structure in the transmission amplitude is due to two effects: (i) The resonance denominator which gives rise to poles, and (ii) the possibility of destructive interference between the first and the second terms which may result in transmission zeros. Note that in contrast to Eq. (1), zeros are now possible.

Next, we examine in more detail conditions for the existence of transmission zeros. As seen from Eq. (5), zeros in the transmission amplitude, $\mathcal{T}_{RL}=0$, occur if

$$\lambda = r_S - \frac{t_{RS}t_{SL}}{t_{RL}} \quad (\text{when } \mathcal{T}_{RL}=0). \quad (7)$$

The above condition relates a property of the resonator, λ , to the characteristics of the junction between the stub and the channel, r 's and t 's. In particular, Eq. (7) can only be true if the right-hand side is a phase factor on the unit circle, i.e., $|r_S - t_{RS}t_{SL}/t_{RL}| = 1$. It is, perhaps, a surprising consequence of unitarity, but easy to show, that always

$$\left| r_S - \frac{t_{RS}t_{SL}}{t_{RL}} \right| = 1. \quad (8)$$

Unitarity, therefore, ensures that both the left-hand side and the right-hand side of Eq. (7) are constrained to the unit circle, which implies that a transmission zero occurs when both phase angles are the same. This proves the existence of transmission zeros for transmission in waveguides.

In the following, we elucidate the above general arguments using specific model systems. We assume that the transmission channels are very thin wires, allowing us to consider only the coordinate along the wire. Transmission and reflection coefficients of such quasi-one-dimensional models have been studied in literature.^{10,11,14} We match the values of the wave functions and the appropriate derivative boundary conditions at each branch point in the network.¹⁵ For the symmetrical open branch, this results in reflection and transmission amplitudes which are independent of energy and orientation, $r_L=r_R=r_S=-1/3$ and $t_{LR}=t_{LS}=t_{SR}=2/3$.

The sidearm can now be made into a resonator of length L by erecting an infinite potential barrier. This forces the wave function to be zero there, and implies $\lambda(E) = -\exp(-2ikL)$. Thus, λ moves around the unit circle with an angular frequency proportional to the wave number. Using Eq (5), the transmission amplitude for this case of a strongly coupled stub can be given analytically, $\mathcal{T} = 2/[2 + i \cot(kL)]$. Figure 3 shows $\mathcal{T}(z)$ in the complex-energy plane: for a complex argument z . Note the appear-

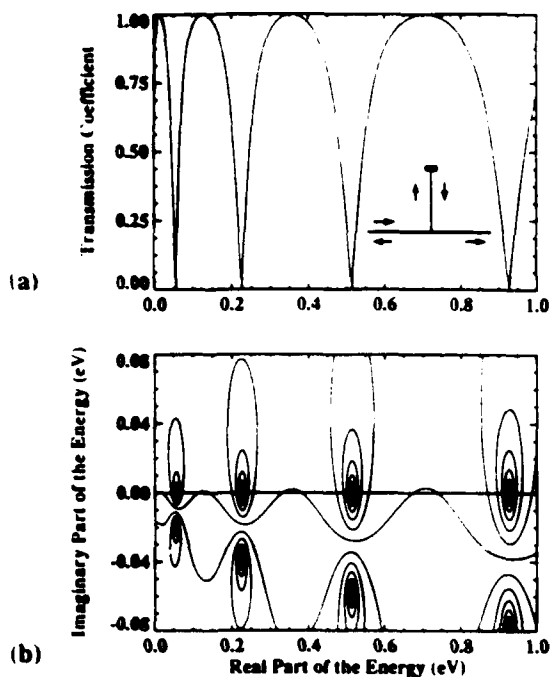


FIG. 3. Transmission amplitude for the strongly coupled stub, which is schematically depicted in the inset: (a) shows the transmission coefficient on the real-energy axis, and (b) shows a contour plot of the absolute value of the transmission amplitude in the complex-energy plane.

ance of transmission zeros on the real-energy axis, and the existence of poles in the fourth quadrant of the complex-energy plane. The zeros occur at energies for which a standing wave forms in the stub, i.e., when $k = n \cdot \pi/L$ with $n = 1, 2, \dots$. Also note that the maxima of the transmission coefficient do not occur at the location of the poles, as for double-barrier resonant tunneling. This is a particularly simple example since the elements of the scattering matrix are independent of energy.

An energy dependence in the elements of the scattering matrix may be introduced by weakly connecting the stub to the channel via a tunneling barrier, as schematically depicted in the inset of Fig. 4. We present numerical results for a tunneling barrier of 0.5-eV height and 1-nm thickness. Figure 4 shows the transmission amplitude for this weakly coupled stub in the complex energy plane. Note again the existence of transmission zeros as predicted by Eq. (8). The poles move closer to the real-energy axis which corresponds to the longer lifetime of the resonant states due to the confining barrier. The most striking feature is the occurrence of a zero-pole pair for each quasi-bound state. With increasing barrier height, the pole approaches the zero which leads to a sharper and sharper transition between a transmission zero and one on the real-energy axis (compare Figs. 3 and 4). In the limit of an infinitely high barrier, the poles and zeros merge which corresponds to $\mathcal{T} = 1$ for a completely decoupled stub.

In summary, we have demonstrated that quasibound states in cavities, which are resonantly coupled to quantum

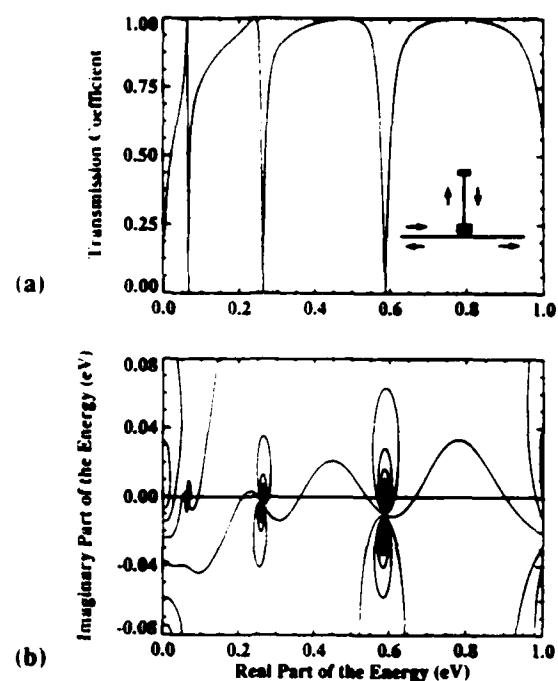


FIG. 4. Transmission amplitude for the weakly coupled stub, which is schematically depicted in the inset: (a) shows the transmission coefficient on the real-energy axis, and (b) shows a contour plot of the absolute value of the transmission amplitude in the complex-energy plane.

waveguides, lead to zero-pole pairs of the transmission amplitude in the complex-energy plane. The proximity of these zeros and poles leads to sharp variations of the transmission coefficient with energy. Zeros are a new "animal" not observed in the usual one-dimensional transmission resonances.

We are grateful for stimulating discussions with Dr. S. Bandyopadhyay, Dr. M. Büttiker, Dr. S. M. Goodnick, Dr. W. Pötz, Dr. P. Price, and Dr. M. Sain. This work was supported, in part, by AFOSR and ONR.

¹ *Nanostructure Physics and Fabrication*, edited by M. Reed and W. P. Kirk (Academic, Boston, 1989).

² *Analogies in Optics and Micro Electronics*, edited by W. Van Haeringen and D. Lenstra (Kluwer Academic, Dordrecht, 1990).

³ F. Sols, M. Macucci, U. Ravaioli, and K. Hess, *J. Appl. Phys.* **66**, 3892 (1989).

⁴ C. S. Lent, in *Computational Electronics*, edited by K. Hess, J. P. Leburton, and U. Ravaioli (Kluwer, Boston, 1990), p. 259.

⁵ P. J. Price, *IEEE Trans. Electron. Devices* **39**, 520 (1992).

⁶ P. J. Price, *Phys. Rev. B* **38**, 1994 (1988).

⁷ M. Büttiker, *IBM J. Res. Dev.* **32**, 63 (1988).

⁸ T. B. Bahder, C. A. Morrison, and J. D. Bruno, *Appl. Phys. Lett.* **51**, 1089 (1987).

⁹ S. Datta, *Quantum Phenomena* (Addison-Wesley, Reading, MA, 1989).

¹⁰ M. Büttiker, Y. Imry, and M. Ya. Azbel, *Phys. Rev. A* **30**, 1982 (1984).

¹¹ M. Cahay, S. Bandyopadhyay, and H. L. Grubin, *Phys. Rev. B* **39**, 12 989 (1989).

¹² D. Kowal, U. Sivan, O. Entin-Wohlman, and Y. Imry, *Phys. Rev. B* **42**, 9009 (1990).

¹³ B. Shapiro, *Phys. Rev. Lett.* **50**, 747 (1983).

¹⁴ C. H. Wu and G. Mahler, *Phys. Rev. B* **43**, 5012 (1991).

¹⁵ J. E. Avron, A. Raveh, and B. Zur, *Rev. Mod. Phys.* **60**, 873 (1988).

Lateral p - n junctions and quantum wires formed by quasi two-dimensional electron and hole systems at corrugated GaAs/AlGaAs interfaces

Wolfgang Porod and Henry K. Harbury

Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

Stephen M. Goodnick

Department of Electrical and Computer Engineering, Oregon State University, Corvallis, Oregon 97331

(Received 13 April 1992; accepted for publication 4 August 1992)

We report the results of modeling lateral p - n junctions and p - n - p quantum wire structures at corrugated GaAs/AlGaAs interfaces, using the surface orientation dependent amphoteric nature of Si doping. We determine the potential landscape and the electron and hole charge densities within a semiclassical Thomas-Fermi screening model, and then solve the two-dimensional Schrödinger equation using finite elements for the quantized electron and hole states at the heterointerfaces. We demonstrate the formation of a one-dimensional electron system confined between two lateral p - n junctions, and discuss the advantages of this structure compared to conventional electrostatic confinement schemes for fabricating quantum wires.

Recent studies¹⁻⁴ have shown that Si acts as an amphoteric dopant depending on substrate orientation during molecular beam epitaxial growth of GaAs and AlGaAs. In particular, growth on {100} surfaces invariably leads to n -type doping, while p -type behavior has been reported for Ga-terminated {111}A surfaces. These findings have led to the investigation of lateral p - n junctions at selectively etched and epitaxially regrown V grooves consisting of {111}A facets on {100} surfaces. Miller³ demonstrated the feasibility of lateral GaAs p - n junctions which exhibited good diode current-voltage characteristics. By using Si-doped $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ rather than GaAs during epitaxial regrowth, one may also realize a lateral p - n junction between quasi-two-dimensional electron and hole systems at the corrugated GaAs/AlGaAs interface. Ebner *et al.*⁵ demonstrated electroluminescence corresponding to the GaAs quantum well band gap in such a system, and Harbury *et al.*⁶ reported calculations which confirmed the existence of a lateral p - n junction between two-dimensional electron and hole gas systems for Si doping densities in excess of $5 \times 10^{17} \text{ cm}^{-3}$.

The ability to fabricate lateral p - n junctions between high mobility two-dimensional hole and electron gases allows for a variety of novel complementary device structures. In previous studies, V-grooves have been employed for the design of quantum wires.⁷⁻¹⁰ By utilizing amphoteric Si in such a V-groove structure, a novel quantum wire system may be realized at lateral p - n - p junctions.¹¹ This new way of fabricating a quantum wire with lateral p - n junctions¹² may offer some advantages compared to present unipolar structures¹³ which use metal gate electrostatic confinement or sidewall etching.

In this letter, we report our results of modeling both lateral p - n junctions and lateral p - n - p quantum wire structures at corrugated GaAs/AlGaAs interfaces. We determine the potential landscape and the electron and hole charge densities by solving Poisson's equation within a semiclassical Thomas-Fermi screening model. Given a certain potential profile, we then solve Schrödinger's equation for the quantized states at the heterointerfaces. Among the

interesting properties of such structures are the coexistence of quasi-two-dimensional electron and hole states next to a quasi-one-dimensional electron or hole system.

Figure 1 shows a schematic drawing of the model corrugated GaAs/AlGaAs interface. The composition and layer thicknesses are similar to those used by Ebner *et al.*⁵ in fabricating a lateral light emitting structure. A 50 nm thick layer of Si-doped $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ is overgrown on a semi-insulating GaAs substrate with an etched V-groove. The substrate terminates in a (100) plane, whereas the sidewalls of the V-groove are members of the {111}A family of planes. As indicated in Fig. 1, n -type doping is present for the (100) layers, and p -type doping for the (111) layers.¹⁴ Also shown is the heterointerface between GaAs and AlGaAs. For suitable doping conditions, a two-dimensional electron and hole gas (2DEG and 2DHG, respectively) forms along this interface.

We use the finite element method and Newton-Raphson iteration to solve the nonlinear Poisson equation within a Thomas-Fermi screening model for room temperature. We assume surface pinning at the $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ (100) and (111) facets (0.8 eV separation between the conduction band edge and the Fermi level), and n -type background doping in the bulk of $1 \times 10^{15} \text{ cm}^{-3}$. The determination of the two-dimensional conduction and valence band edges is a challenging numerical problem which requires a highly nonuniform mesh because of the complicated geometry and the different doping densities. The resulting large system of linear equations is solved utilizing band width optimized sparse matrix methods. We also study the quantum-confined electronic states at the heterointerfaces by obtaining solutions of the two-dimensional Schrödinger equation for the calculated potential landscape.

We now concentrate on the portion of our model structure which is shaded and labeled " p - n Junction" in the schematic drawing of Fig. 1. Depending upon the doping densities in the overgrown AlGaAs layer, the (100) and (111) heterointerfaces may induce two-dimensional electron and/or hole systems, respectively. A lateral p - n junc-

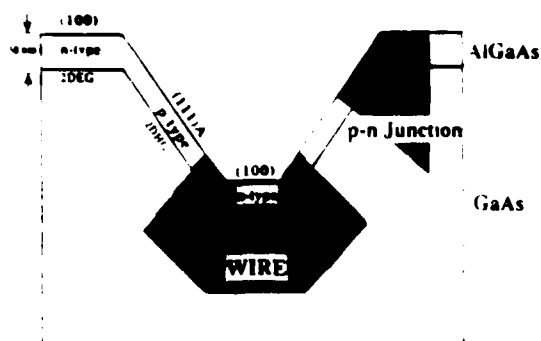


FIG. 1. Schematic diagram of the model geometry.

tion will form if the Si-doping density is high enough to simultaneously produce two-dimensional electron and hole systems along the heterointerfaces. Figure 2 shows, for several values of the Si-doping density in the overlayer, the calculated conduction band profiles at the *n*-side of the *p-n* junction in the direction *perpendicular* to the (100) heterointerface; the semiclassical Fermi level defines the zero of energy and is shown by the dashed line. The inset displays the conduction band minimum at the GaAs side of the heterointerface, and a quasi-two-dimensional electron gas forms for doping densities higher than $5 \times 10^{17} \text{ cm}^{-3}$. Similar behavior is found for the valence bands in the direction *perpendicular* to a (111) interface. Figure 3 demonstrates the existence of lateral *p-n* junctions between 2DEGs and 2DHGs for two values of the overlayer doping density. Plotted are the conduction and valence band edges in a direction *parallel* to the (111) and (100) heterointerfaces, respectively, as schematically shown in the inset. Negative values of the distance correspond to the (111) interface, and positive values to the (100) interface. The emergence of a *p-n* junction is clearly visible, and the transition from *p*- to *n*-type behavior occurs over a distance of about 100 nm.

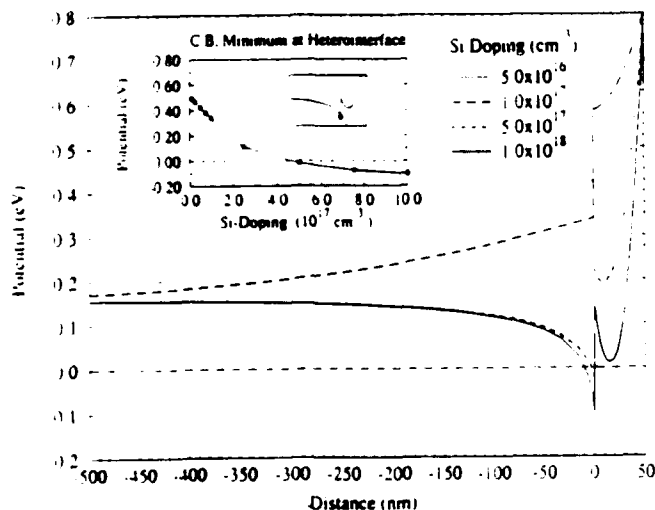


FIG. 2. Conduction band profile *perpendicular* to the (100) interface for different values of the Si-doping density in the overlayer. The inset shows the variation with doping of the conduction band minimum at the heterointerface.

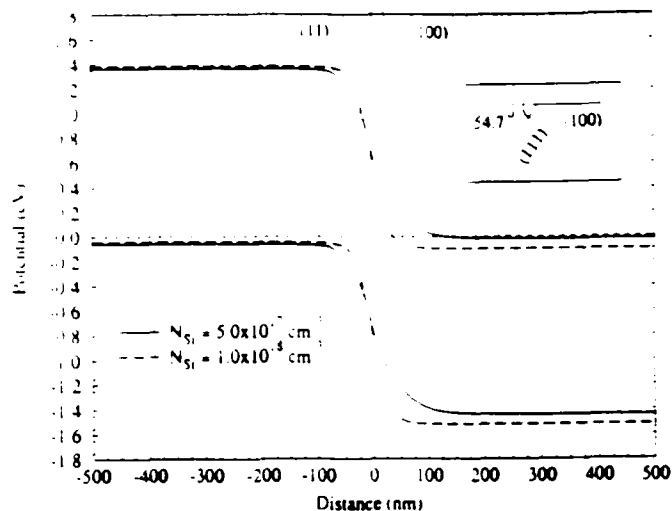


FIG. 3. Band diagram for the lateral *p-n* junction. Shown are the conduction and valence bands (for two values of the doping density) along the (111) and (100) interfaces, as indicated in the inset.

At the base of the V groove, a narrow (100) *n*-type region exists between two (111) *p*-type regions which results in lateral potential confinement in addition to that of the heterojunction. We have investigated the formation of a 1D quantum wire in the *n*-region by solving Poisson's and Schrödinger's equations in the solution domain shown in Fig. 1 by the hatched area which is labeled "WIRE." Figure 4 gives the band diagram for the *p-n-p* junction, where the center *n*-type section is modeled with a length of $L = 100 \text{ nm}$ and the overlayer doping is chosen to be $1 \times 10^{18} \text{ cm}^{-3}$. The conduction and valence band edges are shown at the GaAs side of the junction and in a direction *parallel* to the heterointerfaces. Note that an accumulation of electrons occurs at the center of the *n*-type (100) section where the conduction band moves below the semiclassical Fermi level, which is again chosen to be the zero of energy

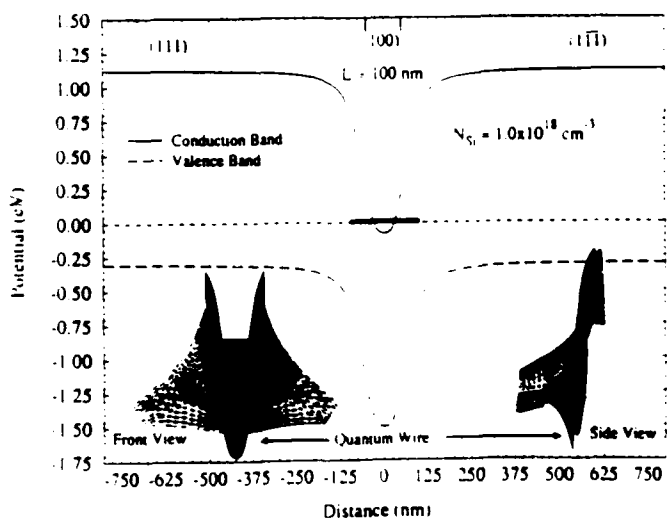


FIG. 4. Band diagram for the lateral *p-n-p* structure. Shown are the conduction and valence bands along the heterointerfaces for a 100 nm long center *n*-type section and a Si-doping density of $1 \times 10^{18} \text{ cm}^{-3}$. The insets show front and side views of the two-dimensional conduction band profile. The potential "pocket" which holds the quantum wire is also indicated.

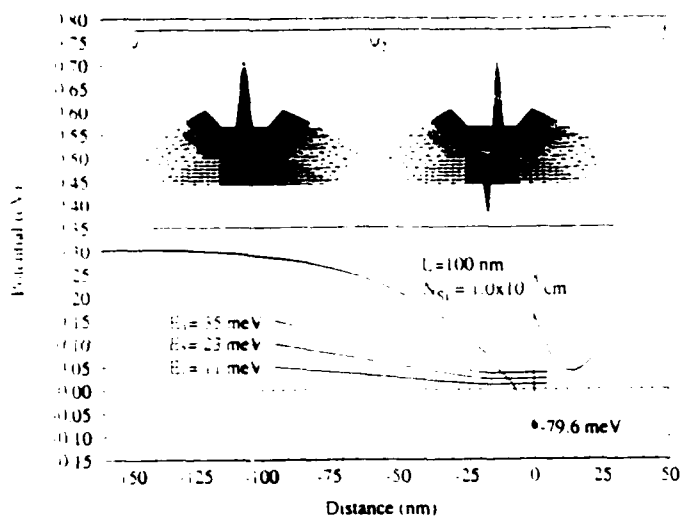


FIG. 5. Potential variation *perpendicular* to the heterointerface (at the center of the *n*-type region) for the two-dimensional conduction band profile shown in Fig. 4. The insets show the wave functions for the lowest two quantum wire states, which are confined at the GaAs side of the heterointerface (the darker top portion of the mesh corresponds to the higher-density AlGaAs overlayer).

and shown as the dashed line. The lateral *p-n* junctions on both sides confine the accumulated electrons in the direction parallel to the heterointerface. A better perspective of the two-dimensional potential variation can be gained by the insets which present front and side views of the two-dimensional conduction band profile. Note the dip in the center which defines the "pocket" holding the electrons.

Figure 5 shows the variation of the conduction band in the direction *perpendicular* to the heterointerface at the center of the *n*-type (100) region. The sharp dip below the semiclassical Fermi level is now clearly visible. The corresponding solutions to the two-dimensional Schrödinger equation for the ground and first excited one-dimensional subbands are shown by the insets in Fig. 5. The peak of the envelope function clearly lies in the GaAs side of the heterojunction localized between the two *p*-type regions. (The darker top portion corresponds to the denser mesh in the AlGaAs overlayer.) The confinement energies relative to the semiclassical Fermi energy are shown schematically on the band diagram for the first three levels. The energies shown all correspond to states originating from the lowest two-dimensional subband energy of the heterojunction itself. The spacing of the levels, here on the order of 10 meV, depends on the doping and the width of the *n*-region at the base of the V-groove. As the width is reduced, the subband separation should increase, although limitations are imposed by the lateral extent of the depletion region which eventually results in complete depletion of the wire if made too narrow.

The results above demonstrate the feasibility of realizing a quantum wire structure on a corrugated GaAs/AlGaAs surface. There are several possible advantages of such a structure compared to the current state of the art which relies on sidewall etching or electrostatic confinement from Schottky contacts on the surface of the AlGaAs. First, the actual definition of the width of the

n-region is defined by the etch time of the anisotropic etch through an optically defined photoresist mask. Thus, electron or ion-beam lithography is not required to fabricate the wire structure. The bipolar nature of the structure may be utilized to inject minority carriers into the wire structure from the *p*-regions, which may lead to some interesting device applications. Also, if one additionally provided separate contacts to the *p*- and *n*-type regions using nanolithography, control of the wire transport properties could be achieved both by varying the width of the confinement (through reverse biasing the *p-n* junctions), and through varying the electron Fermi energy via a gate contact on the AlGaAs above the electron channel.

In summary, we have demonstrated the existence of lateral *p-n* junctions between 2DEGs and 2DHGs on corrugated GaAs/AlGaAs interfaces, which is made possible by the amphoteric nature of Si-dopants as a function of crystallographic plane orientations. At V-grooves etched into a semi-insulating GaAs substrate, a *p-n* junction forms at the intersection of (100) and (111) planes if the doping in the overlayer exceeds $5 \times 10^{17} \text{ cm}^{-3}$. We have also demonstrated the possibility of basing this system for the design of quantum wires which may form at the bottom of the V groove. We find that a quantum wire is to be expected in this *p-n-p* structure for a length of the *n*-type region on the order of 100 nm and a Si-doping density of $1 \times 10^{18} \text{ cm}^{-3}$.

The authors would like to acknowledge fruitful discussions with J. Ebner, D. J. Kirner, C. S. Lent, J. L. Merz, M. Mueller, and T. K. Plant. This work has been partially supported by the Office of Naval Research and the Air Force Office of Scientific Research.

- ¹J. M. Ballingall and C. E. C. Wood, *Appl. Phys. Lett.* **41**, 947 (1982).
- ²W. I. Wang, E. E. Mendez, T. S. Kuan, and L. Esaki, *Appl. Phys. Lett.* **47**, 826 (1985).
- ³D. L. Miller, *Appl. Phys. Lett.* **47**, 1309 (1985).
- ⁴S. Subbanna, H. Kroemer, and J. L. Merz, *J. Appl. Phys.* **59**, 488 (1986).
- ⁵J. Ebner, J. E. Lary, G. W. Eliason, and T. K. Plant, *Proceedings of the 20th European Solid State Device Research Conference, ESSDERC 90*, edited by W. Eccleston and P. J. Rossner (Adam Hilger, Bristol, 1990), pp. 401-404.
- ⁶H. K. Harbury, W. Porod, and S. M. Goodnick, presented at the 19th International Conference on the Physics and Chemistry of Semiconductor Interfaces, Death Valley, CA, January 1992.
- ⁷B. I. Miller, A. Shahar, U. Koren, and P. J. Corvini, *Appl. Phys. Lett.* **54**, 188 (1989).
- ⁸E. Kapon, D. M. Hwang, and R. Bhat, *Phys. Rev. Lett.* **63**, 430 (1989); and E. Kapon, S. Simhony, R. Bhat, and D. M. Hwang, *Appl. Phys. Lett.* **55**, 2715 (1989).
- ⁹E. Colas, S. Simhony, E. Kapon, R. Bhat, D. M. Hwang, and P. S. D. Lin, *Appl. Phys. Lett.* **57**, 914 (1990).
- ¹⁰H. K. Harbury and W. Porod, *J. Vac. Sci. Technol. B* **8**, 923 (1990).
- ¹¹J. Ebner (private communication).
- ¹²The use of *p-n* junctions for lateral confinement of quantum wires was pioneered by C. C. Dean and M. Pepper, *J. Phys. C* **15**, L1287 (1982), and by A. B. Fowler, A. Hartstein, and R. A. Webb, *Phys. Rev. Lett.* **48**, 196 (1982).
- ¹³*Nanotechnology Physics and Fabrication*, M. A. Reed and W. P. Kirk, Eds. (Academic, San Diego, 1989).
- ¹⁴The details of the amphoteric nature of the silicon doping near the intersection of the (100) and (111) surfaces are not known, and likely to be more complicated than our model. This and possible compensation effects might tend to reduce the abruptness of the lateral *p-n* junction.



Stability of Thin Resist Walls

Xiaokang Huang, Greg Bazán, Davide A. Hill,* and Gary H. Bernstein

Department of Electrical Engineering and Department of Chemical Engineering,
University of Notre Dame, Notre Dame, Indiana 46556*

ABSTRACT

The physical strength of poly (methyl methacrylate) (PMMA) resist is an important factor for making very dense patterns with most types of lithography. For lines placed closely together using the lift-off technique, a thin wall of PMMA remains to act as the metal line spacer. When placed in contact with a suitable developer, thin polymer walls may swell and buckle, depending on the desired aspect ratio of the thin walls. We use electron beam lithography and a high-contrast developer solution to study the relationship of the maximum height-to-width ratio of the PMMA walls necessary to avoid buckling, and the resulting effects on metal thickness after lift-off. Maximum achievable aspect-ratios for our developer system were found to be constant (nearly 5:1) for widths from 20 to 45 nm and lengths of tens of microns. Theoretical predictions based on the hypothesis of a swelling-induced elastic instability are in excellent agreement with the experiments.

In the past few years, quantum devices based on ultra-small geometries have received considerable attention. The usual case involves some feature, *e.g.*, gate length or metal line width, which is reduced in size in order to observe size quantization. In some cases, reduction of line widths plays only half the role, whereas a decrease in spacing between features plays an equally significant part. For example, split gates with 100 nm gaps in the form of bends have been made for the study of transmission properties at corners.¹ Also, lateral resonant tunneling transistors have been

demonstrated with distances between gates of about 100 nm.² More recently, optical gratings have been fabricated with a metal line pitch of 50 nm over several microns.³ All such devices require excellent control of line spacing.

Because of its uniquely high resolution, poly (methyl methacrylate) (PMMA) is among the most important resists for nanostructure fabrication. Utilizing the lift-off process, the resolution of resulting metal patterns depends not only on the width of the defined lines, which relates to the resolution of the beam formation system, but also to the quality of the edge profile of the resist, which depends to a great extent on the contrast of the resist/developer system. The

* Electrochemical Society Active Member.

most common developers for nanolithography are methyl isobutyl ketone:isopropyl alcohol (MIBK:IPA) (1:3)¹ and cellosolve:methanol (3:7),² both demonstrating very high resolution. One weakness of these developers is the inability to dissolve high molecular weight components left behind from doses just at the threshold of complete development. This can lead to scum layers which can interfere with the lift-off process.³ Higher doses necessary to decrease the molecular weight of the scum layer lead to a loss of resolution due to such factors as beam tails and energy scattering. It has been shown⁴ that adding a minute percentage of methyl ethyl ketone (MEK) to the common developers can increase contrast and improve lift-off results, and therefore allow denser patterns to be fabricated.

In this paper we discuss the fabrication of extremely dense patterns on a solid substrate at energies below 50 keV without the use of such techniques as shadowing of a single pass exposure⁵ or re-alignment of separate patterns.⁶ Although patterns as small as 15 nm have been achieved with double tilt shadow evaporation,⁷ a major difficulty is that of connecting the resulting lines to other structures within a device. Double alignment techniques⁸ have shown equally impressive results with 16 nm lines on 24 nm pitch, but this method is inherently difficult and complicated, relies on the use of a scanning transmission electron microscope (STEM) for aligning the patterns, and requires multiple exposure steps. So far, only membranes, as opposed to solid substrates, have been utilized in this technique since alignments are facilitated using transmission imaging. Although adding significant complexity to the process, membranes offer an inherent advantage of reduced proximity effects, and thus a higher exposure modulation between the lines and spaces. Unfortunately, however, membranes are not amenable to device fabrication or suitable device performance. The developer used here helps to improve contrast so that very dense patterns can be fabricated on solid substrates even without very high beam energies offered by the use of TEMs.

Both the beam diameter of the exposure tool and the contrast of the resist/developer system have a significant effect on the achievable density of narrow metal lines. In the limit of very high resolution and contrast, the mechanical strength of the PMMA used for lift-off is the final limitation to making extremely closely spaced patterns. The lift-off process is described in Fig. 1. Exposing energy, in our case electrons, causes bond-breaking in the PMMA (Fig. 1a). After development (Fig. 1b), metal is evaporated perpendicularly to the surface, so that metal coats only the top of the resist and the exposed substrate surface (Fig. 1c). For thin resists at moderate energies, undercutting can approximately be described by⁹

$$b = 625 (Z/E_e) t^{1/2} (\rho/A)^{1/2} \quad [1]$$

where b is the increase of the diameter of the beam at the bottom surface of the resist, Z , A , ρ , and t are the atomic number, atomic weight, density, and thickness of the resist, respectively, and E_e is the beam energy in keV. For PMMA, $Z = 3.6$, $A = 6.7$ g/mol, and $\rho = 1.2$ g/cm³.¹⁰ For 45 keV electrons and a PMMA thickness of 130 nm, undercutting is about 5 nm at each sidewall of our pattern, which aids in the lift-off process. During the lift-off step, dissolution of the resist in a suitable solvent causes removal of all metal except that on the substrate surface (Fig. 1d).

As seen above, the unexposed resist spacer is in the form of a thin wall between the developed lines. When lines are placed too closely together and high resolution is achieved, the resist walls can either become wavy, fall over, or even lose adhesion to the surface, in which case they fail completely. In this regard, internal stresses due to swelling can play a dominant role. Although PMMA suffers little swelling, which accounts partially for its ultrahigh resolution properties, some absorption of developer is necessary to aid the development process.¹¹ Swelling of polymeric resists has been extensively investigated.¹² In poor solvents, solvation forces favoring mixing are insufficient to promote dissolution of the polymer. Partial plasticization and

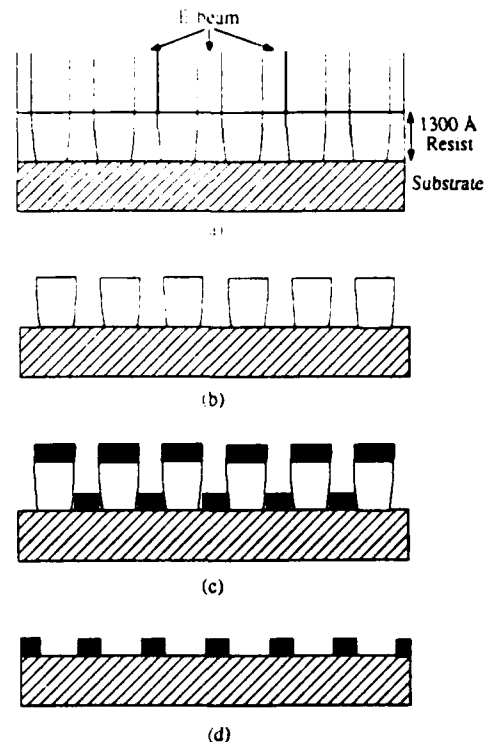


Fig. 1. Explanation of the lift-off process. (a) Exposure of resist by electron beam before development, resulting in slight undercutting. (b) Resist wall profile after development. (c) Resist walls after metalization from small, perpendicular source. (d) Metal pattern after dissolution of remaining resist in suitable solvent.

swelling can occur, however, leading to weaker mechanical properties, and to the build up of internal stresses in constrained systems. Typically, glassy polymers can experience between 1 and 30% swelling, without dissolution, depending on solvent quality and temperature.¹³ Such a low degree of expansion may be irrelevant in most cases, but, as shown below, can lead to severe dimensional stability problems in the fabrication of thin polymer walls.

Experimental

Our EBL system consists of an Amray 1400 scanning electron microscope (SEM) with a minimum spot size of 35 Å controlled by an IBM PS/2 personal computer interfaced through a Pragmatic Systems 2201A 16-bit arbitrary waveform generator. The details of the system are described elsewhere.¹⁴

We used PMMA with molecular weight 950,000 amu baked at 170°C for 4 h for all exposures. Resist and developer parameters were very tightly controlled. Resist thickness was 105 to 180 nm as measured with an Alpha-Step 200 surface profiler. The developer used in all cases was a solution of MIBK:IPA (1:3) with the addition of 1% MEK by volume.⁹ It has been shown that contrast results are extremely sensitive to variations in developer temperature due to the high activation energy of the PMMA in the developer solvents.⁶ Using a temperature controller, our developer temperature was maintained at $25 \pm 0.5^\circ\text{C}$. Develop time for all samples was 15 s followed by a 15 s rinse in methanol. After development, all samples were dried in air at 90°C for 5 min.

All exposures were performed at a beam energy of 45 keV at a spot size of 5 nm and 120 μm scan field. Substrates used in all cases were 200 nm SiO₂ on thick silicon wafers. Doses were very critical for making high-density patterns because of proximity effects during E-beam lithography. They were not only related to the line pitch, but also to the size of the patterns. For grating pitches ranging from 70 to 150 nm, the doses varied from 1.5 to 2.5 nC/cm. For the case of only two parallel lines, doses ranged from 2.5 to 3.5 nC/cm for pitches from 80 to 120 nm.

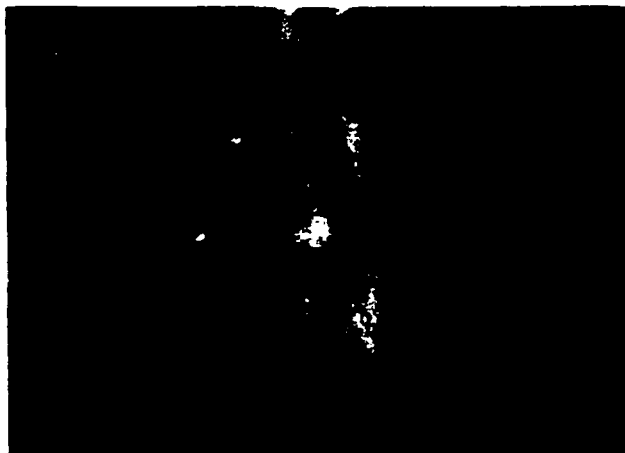


Fig. 2. Remaining metal after lift-off of a typical failed wall. Light areas are gold, and dark, wavy line is the shadow of failed PMMA wall.

All metal evaporations were gold from an electron beam evaporator with crucible size of 2.5 cm and a distance of 40 cm from the source to wafer. Metal thicknesses were closely calibrated with a film thickness monitor and a surface profilometer. All metal thicknesses were 30 ± 1 nm. Lift-off was performed by soaking the samples in acetone for about 5 min, followed by the application of acetone using a hypodermic syringe, followed by a 15 s methanol rinse, and final blow drying in nitrogen.

Results

In order to achieve lift-off of evaporated metal films, it is necessary that PMMA walls remain between the developed areas. Assuming minor undercutting for thin resists and high beam energies, it is expected that the spaces of the resulting metal patterns closely reflect the width of the remaining PMMA walls, especially at the top. Because of the very low solubility of unexposed PMMA of very high molecular weight,⁶ the height of the walls after development is still approximately the same as the initial resist thickness.

After application and bake, PMMA can be considered a strong, glassy material. Exposure to the developer, however, causes swelling of the polymer, leading to the buildup of internal stresses in strongly anchored structures. Compressive stresses in the wall can, in turn, lead to buckling.

The wall aspect-ratio (height-to-width) is an important parameter in delineating the stability threshold, since a tall, thin wall will be strongly susceptible to small dimensional perturbations around the erect position, while a short wall will be able to sustain a higher degree of swelling before buckling. Figure 2 shows the remaining metal after lift-off of a typical failed wall. The bright areas are evaporated gold, and the dark, wavy center line is the shadow of the remaining wall over which gold was deposited and removed. The wall was about 30 nm wide and 140 nm high. We show later that a direct relationship exists between the original wall height and the wavelength of the distortion.

Figure 3 shows the minimum width of a PMMA wall achievable for a given starting resist thickness. Error bars represent variations in metal line separation (and therefore wall thickness) over the length of the lines. These data were obtained by smoothly varying the line pitch for a given resist thickness and observing the point at which lines either failed completely or became noticeably wavy. The data indicate a direct proportionality between critical wall height and width, implying a constant aspect-ratio for buckling.

Although the lower the thickness of the PMMA, the closer the achievable line spacing, the resist thickness cannot be lowered arbitrarily because, for different processes, different thickness of metal are required depending on step coverage, line continuity, multiple metal layers, device scaling requirements, electrical resistance constraints,

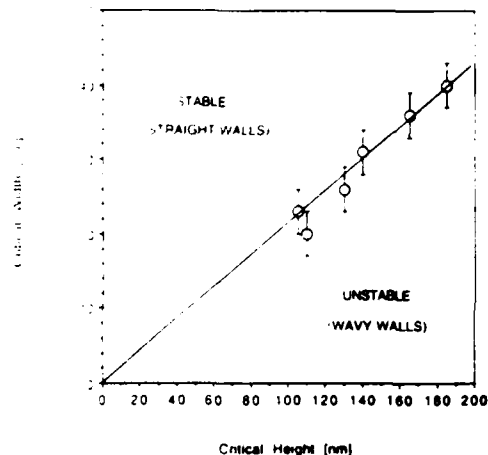


Fig. 3. Minimum width of PMMA walls achievable for a given starting resist thickness (wall height).

opacity to radiation (e.g., x-rays), etc. We have determined that to achieve reliable lift-off, the thickness of PMMA should be at least three times the desired thickness of the small metal pattern. This rule of thumb will depend to some extent on the metal evaporation system, since the physical size of the evaporant source, the source-to-sample spacing, and the care taken to place the sample perpendicular to and directly above the source will all affect the amount of metal deposited on the inside surfaces of the PMMA etched by the developer. Using this relationship, and guided by the "critical-aspect-ratio" criterion, we fabricated double gold lines with an approximately 20 nm space and a metal thickness of 30 nm, from 105 nm thick resist layers. The lines were continuous and straight over a distance of 25 μ m. Figure 4 shows a portion of these lines at high magnification.

Discussion

We begin our discussion with some simple theoretical considerations about the stability of a long wall with a rigidly anchored bottom and a top free end subject to internal stresses. Comparison with experiments will then closely follow the theoretical predictions.

The system of interest is shown in Fig. 5, which also gives the characteristic geometrical parameters. Mathematically, the problem is conveniently formulated within the framework of linear plate theory,¹⁵ leading to the following homogeneous boundary value problem for the lateral displacement, δ

$$\frac{\partial^4 \delta}{\partial x^4} + 2 \frac{\partial^2 \delta}{\partial x^2 \partial y^2} + \frac{\partial^2 \delta}{\partial y^4} = - \left(\frac{\sigma_{xx} 12(1-\nu^2)}{Ew^3} \right) \frac{\partial^2 \delta}{\partial x^2} \quad [2]$$



Fig. 4. Very closely spaced gold lines at high magnification. The distance between lines is 20 nm over a distance of 25 μ m.

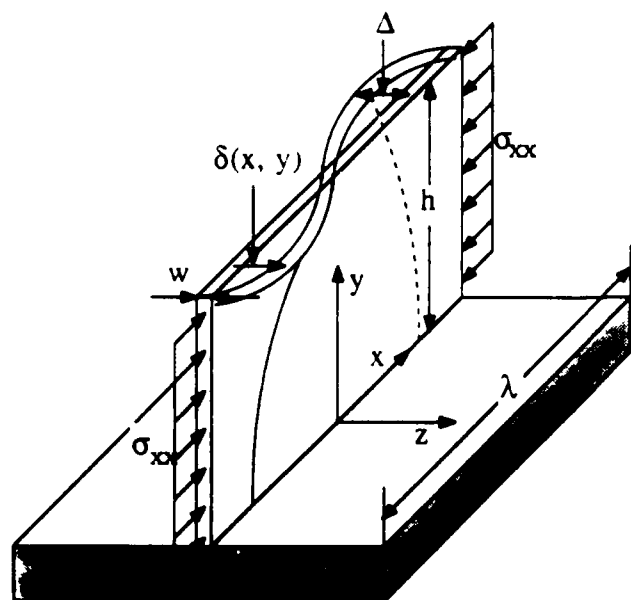


Fig. 5. Diagram of tall, thin wall demonstrating geometrical parameters used in the analysis.

$$\delta = 0 \text{ at } y = 0 \quad (\text{no displacement at the bottom}) \quad (3)$$

$$\frac{\partial \delta}{\partial y} = 0 \text{ at } y = 0 \quad (\text{built-in bottom end}) \quad (4)$$

$$\frac{\partial^2 \delta}{\partial y^2} + \nu \frac{\partial^2 \delta}{\partial x^2} = 0 \text{ at } y = h \quad (\text{torque-free top}) \quad (5)$$

$$\frac{\partial^2 \delta}{\partial y^2} + (2 - \nu) \frac{\partial^2 \delta}{\partial x^2} = 0 \text{ at } y = h \quad (\text{shear force-free top}) \quad (6)$$

where $\sigma_{xx} \geq 0$ is the internal compressive stress, ν Poisson's modulus, E Young's modulus, and w and h the wall thickness and height, respectively. For an infinitely long wall the solution to Eq. 2 is of the form

$$\delta(x, y) = \cos(\xi x/h) [A(\sinh(\alpha y/h) - \sin(\beta y/h)) + B(\cosh(\alpha y/h) - \cos(\beta y/h))] \quad (7)$$

where A and B are integration constants, and

$$\alpha = \sqrt{\xi^2 + \nu \Lambda^2 \xi^2} \quad (8)$$

$$\beta = \sqrt{-\xi^2 + \nu \Lambda^2 \xi^2}$$

$$\Lambda = \frac{\sigma_{xx} 12(1 - \nu^2) h^2}{E w^3}$$

Variable ξ is related to the wavelength of the prevailing distortion, λ , as follows

$$\lambda = 2\pi h / \xi \quad (9)$$

Imposition of the boundary conditions provides the following equation for the eigenvalues

$$2ts + (s^2 + t^2) \cosh(\alpha) \cos(\beta) = \frac{1}{\alpha\beta} (\alpha^2 t^2 - \beta^2 s^2) \sinh(\alpha) \sin(\beta) \quad (10)$$

where

$$t = \beta^2 + \nu \xi^2 \quad (11)$$

$$s = \alpha^2 - \nu \xi^2$$

As customary in stability theory, Eq. 10 is satisfied only by a discrete set of ξ values, provided Λ is greater than or equal to a minimum critical value, Λ^* . Physically, Λ^* is the smallest (dimensionless) load necessary to reach the mar-

ginal stability limit (indifferent equilibrium), and can be calculated by solving Eq. 10 numerically. We obtain $\Lambda^* = 10.40$, and $\xi^* = 1.917$. Combining these results with Eq. 8 and 9, the smallest critical stress, σ_{xx}^* , and the associated wavelength, λ , are calculated to be

$$\sigma_{xx}^* = 26 E w^2 / [30(1 - \nu^2) h^2] \quad (12)$$

$$\lambda = h 2\pi / 1.917 \quad (13)$$

For swelling-induced stresses, σ_{xx}^* can in turn be obtained from

$$\sigma_{xx}^* = E \epsilon_{sw} \quad (14)$$

where ϵ_{sw} is the hypothetical swelling strain that the wall would undergo if allowed to expand freely. Equations 12 and 14 allow us to calculate the critical wall aspect-ratio for stability, $(h/w)^*$, as a function of the swelling strain

$$\left(\frac{h}{w}\right)^* = \sqrt{\frac{26}{30(1 - \nu^2)\epsilon_{sw}}} \quad (15)$$

For $(h/w) > (h/w)^*$ the wall will buckle. Equation 15 is more convenient for comparison with experiments than Eq. 12, since it does not contain the unknown modulus E (which presumably changes upon swelling). As expected, a short wall can sustain a higher degree of swelling before buckling. Additionally, Eq. 15 predicts that, for a constant ϵ_{sw} , the critical wall height and width should be directly proportional to each other. This prediction is in excellent agreement with our data, as shown in Fig. 3.

The slope of the line in the figure provides the critical aspect ratio for buckling: $(h/w)^* = 4.75$. Calculation of ϵ_{sw} from Eq. 15 requires knowledge of Poisson's modulus, ν . Because the polymer is swollen, and the stresses are relatively small, the assumption $\nu = 0.50$ is appropriate, giving: $\epsilon_{sw} = 5.1\%$. This value is well within the expected range for glassy polymers.¹³

Another easily accessible observable is the wavelength of the distortion after buckling, λ , (Eq. 13). In Fig. 2, the measured wavelength of the distortion is 460 nm, which is in excellent agreement with the value of 459 nm given by Eq. 13 with $h = 140$ nm.

The maximum amplitude of the lateral displacement of the wall, Δ (see Fig. 5), can also be measured by SEM. If we assume that the deformation profile is unaffected by solvent evaporation, then the observed value of Δ can be used to obtain an independent estimate of ϵ_{sw} as follows.

The curvilinear contour length between nodal points at the top of the wall after buckling can be calculated from

$$L = \lambda(1 + \epsilon_{sw}) = \frac{\lambda}{2\pi} \int_0^{2\pi} \sqrt{1 + \left(\frac{2\pi\Delta}{\lambda}\right)^2 \sin^2(z)} dz \quad (16)$$

For $(2\pi\Delta/\lambda)^2 \ll 1$ (subject to verification), the kernel in Eq. 16 can be expanded up to second-order terms and integrated to give

$$\epsilon_{sw} = \left(\frac{\pi\Delta}{\lambda}\right)^2 \left[1 - \frac{3}{2} \left(\frac{\pi\Delta}{\lambda}\right)^2\right] \quad (17)$$

This equation can be used to calculate ϵ_{sw} , with the measured values of Δ and λ . From Fig. 2 we estimated $\Delta \approx 33$ nm, giving (with $\lambda = 460$ nm) $\epsilon_{sw} = 4.7\%$. This value is in good agreement with the previous estimate of 5.1%. These findings confirm that swelling is the dominant mechanism for instability in our experiments. It is remarkable that a 5% swelling can lead to the failure of a wall less than five times higher than its width.

Summary and Conclusions

We have investigated the relationship between minimum line spacing and maximum metal thickness for EBL-fabricated patterns using high-contrast resists. Both theory and experiments confirm that a small degree of swelling may lead to the catastrophic failure of PMMA walls of relatively modest aspect-ratio. For subcritical aspect-ratios, line spacings as small as 20 nm, and metal thicknesses as high as 30 nm were reliably achieved without loss of adhesion or

distortion. We believe that the formulation of developers that maximize contrast while minimizing swelling is desirable for attaining higher resolution in nanostructure fabrication. Hopefully, the results of this paper will permit more efficient screening of improved developer solutions through the initial observation of swelling properties.

Acknowledgments

We would like to thank D. T. Leighton for helpful discussions and S. Subramaniam for help in sample preparation. This research was supported in part by IBM, AFOSR, ONR, and the University of Notre Dame. D.A.H. gratefully acknowledges the Donors of The Petroleum Research Fund, administered by the American Chemical Society, for partial support of this research.

Manuscript submitted April 30, 1992; revised manuscript received June 22, 1992.

The University of Notre Dame assisted in meeting the publication costs of this article.

REFERENCES

1. J. C. Wu, M. N. Wybourne, W. Yindeepol, A. Weisshaar, and S. M. Goodnick, *Appl. Phys. Lett.*, **59**, 102 (1991).
2. D. R. Allee, S. Y. Chou, J. S. Harris, Jr., and R. F. W. Pease, *J. Vac. Sci. Technol. B*, **7**, 2015 (1989).
3. S. Y. Chou, Y. Liu, and P. B. Fischer, *IEDM Tech. Dig.*, 91, 745 (1991).
4. S. P. Beaumont, P. G. Bower, T. Tamamura, and C. D. W. Wilkinson, *Appl. Phys. Lett.*, **38**, 436 (1981).
5. H. G. Craighead, R. E. Howard, L. D. Jackel, and P. M. Mankiewich, *ibid.*, **42**, 38 (1983).
6. G. H. Bernstein, D. A. Hill, and W. P. Liu, *J. Appl. Phys.*, **71**, 4066 (1992).
7. S. Y. Chou and P. B. Fischer, *J. Vac. Sci. Technol. B*, **8**, 1919 (1990).
8. K. Y. Lee, J. Frost, C. Stanley, W. Patrick, W. S. Mackie, S. P. Beaumont, and C. D. W. Wilkinson, *ibid.*, **5**, 322 (1987).
9. J. I. Goldstein, in *Introduction to Analytical Electron Microscopy*, J. J. Hren, J. I. Goldstein, and D. C. Joy, Editors, p. 101, Plenum Press, New York (1979).
10. G. Owen, *Reps. Prog. Phys.*, **48**, 795 (1985).
11. A. Weill, in *The Physics and Fabrication of Microstructures and Microdevices*, M. J. Kelly and C. Weisbuch, Editors, p. 58, Springer-Verlag, New York (1986).
12. W. M. Moreau, *Semiconductor Lithography*, Chap. 10, Plenum Press, New York (1988).
13. G. Park and K. Ueberreiter, in *Diffusion in Polymers*, J. Crank and G. Park, Editors, pp. 140-162, 219-257, Academic Press, Inc., New York (1968).
14. G. Bazán and G. H. Bernstein, In preparation.
15. S. P. Timoshenko and J. M. Gere, *Theory of Elastic Stability*, 2nd ed., p. 363, McGraw-Hill, Inc., New York (1961).

Determination of Fixed Electron-Beam-Induced Positive Oxide Charge

G.H. BERNSTEIN, S.W. POLCHLOPEK, R. KAMATH, W. POROD

Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana, USA

Summary: Contrary to previous beliefs, electron-beam-induced positive charges in insulators persist where created for at least several weeks without discernible movement. Formerly, coating with a thin metal overlay was thought to allow the charge to leak away. Coating with a conductor is shown to shield electric fields from affecting the imaging probe, but to remove no charge from the specimen. A new technique is introduced for the evaluation of the properties of electron-beam-induced positive charges in metal-oxide-semiconductor (MOS) capacitors. MOS structures were subjected to partial area exposure in a scanning electron microscope. These exposures resulted in the creation of areas of localized positive charge within the oxide, which was observed as steps in the capacitance-voltage data. A systematic study was performed. It related the exposed area to the step height and the amount of induced charge to the voltage shift of the step. A model describing the observed phenomenon is presented, followed by a comparison of theoretical and experimental results. The progress of the charge over time was studied by performing capacitance-voltage analysis 30 min after electron beam exposure and up to 4 weeks later.

Introduction

Charging of specimens is a common problem encountered in the imaging of all insulating materials in a scanning electron microscope (SEM). In general, the details of the charging mechanism can be a complex function of beam and material parameters. Krause and co-workers (Krause *et al.* 1989) have demonstrated that in thin films,

trapped charge can be either positive or negative, depending on the beam-accelerating voltage, the thickness of the insulating layer, and the nature of the substrate. Typically, at higher voltages, SEM images of thin insulators on conducting substrates appear dark because of trapped positive charges left behind by secondary electrons ejected by the primary beam. It is commonly believed that the coating of insulating specimens by a conductive layer causes this positive charge to be drained away from the imaging area, thus eliminating distortions of the probe and restoring usable resolution (Goldstein *et al.* 1981). A nuisance in SEM imaging, excess positive charge in the gate oxide of metal-oxide-semiconductor (MOS) devices (McLean *et al.* 1989, Nicollian and Brews 1982) can be created by electron beam processing used in certain integrated circuit fabrication and analysis techniques. With electron beam lithography, metal over- or under-layers are used to reduce the charging effects during beam writing of nonconductive resists (Henderson 1980). The precise properties and behavior of the induced charges in insulating films are uncertain. Although in MOS structures much is known about the creation, minimization, and annealing of e-beam-induced bulk oxide damage (Balasinski *et al.* 1988, Keery *et al.* 1976, Sah *et al.* 1983, Shimaya *et al.* 1983), details of the behavior of localized positive charge within oxide layers are not fully understood. This is partly because an easy and accurate method to monitor localized oxide charge has not been established.

For the first time we can report that electron-beam-induced positive charges remain fixed where created. To demonstrate this, we have developed a technique that uses a certain structure in the capacitance-voltage (C-V) curves of MOS capacitors. This structure can be used to yield valuable information regarding the location and amount of charge deposited in SiO₂ in a MOS structure. Using electron beam lithography, we accurately exposed partial areas of MOS capacitors with known electron doses. This served to effectively create two capacitors in parallel, with one metal gate, having different threshold values due to the induced positive charge in the oxide. By performing C-V measurements before exposure, 30 min afterward, and later, it was possible to electrically monitor the induced positive charge. The results of these experiments, along with conclusions about the electrical behavior of the charge within the oxide, are reported. A model relating the exposed capacitor area to the theoretical C-V results is proposed and compared with experimental results.

This work was supported in part by ONR, AFOSR, IBM, and the University of Notre Dame.

Address for reprints:

G.H. Bernstein
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556, USA

Experiment

The starting material used for MOS capacitor fabrication was <100> orientation, p-type silicon with doping concentration $N_A = 1.5 \times 10^{15} \text{ cm}^{-3}$. After careful cleaning, the silicon wafers were oxidized at 1000°C for 20 min, resulting in oxide thicknesses ranging from 738–833 Å, as deduced from the accumulation capacitance of the individual C-V curves. In order to minimize the effects of mobile ion contamination, oxygen was bubbled through a solution of 2% HCl in deionized water. Circular aluminum dots of an area $4.9 \times 10^{-3} \text{ cm}^2$ were evaporated in a high vacuum through a shadow mask onto the surface of the oxidized wafers. The thickness of the Al for all of the dots was 1,100 Å as measured with an Alpha-step 200 profilometer. The top surfaces of the metallized wafers were coated with a protective layer of positive photoresist, then immersed in a buffered hydrofluoric acid (HF) chemical etching solution which removed the oxide from the back surface. Following photoresist removal, aluminum was evaporated onto the back surface of the wafers, after which the wafers were annealed in flowing nitrogen gas for 20 min at 420°C .

Partial areas of the completed MOS capacitors were subjected to electron beam exposure as shown schematically in Figure 1a, changing the device properties where exposed and thus effectively dividing the device into two separate capacitors in parallel, as described in Figure 1b. Exposures were performed at 20 kV and 0.8 nanoamperes in an Amray model 1400 scanning electron microscope. The SEM was controlled using a Pragmatic Instruments model 2201A arbitrary waveform generator and custom software to ensure dose and area accuracy. Exposures were executed by quickly imaging a single capacitor at a very low current, blanking the beam, setting the current to the desired value, and scanning (using e-beam lithography) a square of known area entirely within the MOS capacitor. The dose was fixed at $4 \times 10^{-5} \text{ C/cm}^2$, and exposed areas ranged from 10.4–62.4% of the total area of the capacitor. The upper limit on areal exposures was determined by the size of the largest possible inscribed square within a circle (with area ratio $2/\pi$).

Results and Analysis

Total electron beam exposure of high quality MOS devices is known to cause a negative shift of C-V characteristics (Johnson 1975, Keery *et al.* 1976, McLean *et al.* 1989, Nicollian and Brews 1982, Sah *et al.* 1983, Shimaya *et al.* 1983). Free electrons created in the oxide during exposure either recombine with holes or leave the insulator; thus they are electrically eliminated (Keery *et al.* 1976, Zaininger and Holmes-Siedle 1967). Holes, however, behave differently because of their much lower mobility in SiO_2 , as some will recombine with electrons, while a significant fraction is captured in stationary traps, resulting

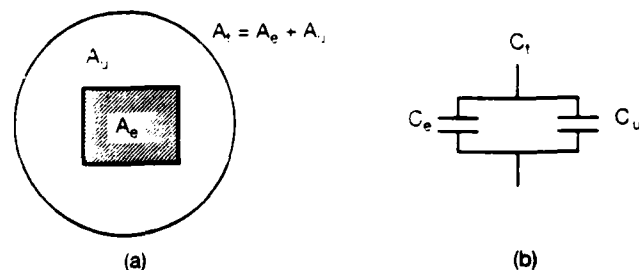


FIG. 1 Schematic diagram of the partial area exposure of a MOS capacitor. (a) Definition of the exposed area, A_e , and the unexposed area A_u . (b) The exposed and unexposed portions give a total capacitance, C_t , which is the sum of the two parallel capacitances, C_e and C_u .

in a fixed positive charge contained in the oxide (Keery *et al.* 1976, Zaininger 1966, Zaininger and Holmes-Siedle 1967). The result of this trapped positive charge is depletion and inversion of the device at more negative voltages.

C-V measurements were performed on a series of MOS capacitors both before and 30 min after partial area exposure in an SEM, as described above. Figure 2 shows typical results obtained from two of these capacitors that have fractional exposure areas, A_e/A_t , where A_e is the exposed area and A_t is the total area, of 10.4 % and 62.4%, respectively. The relevant features in this figure are the additional steps in the C-V curves observed after exposure. Comparison of Figures 2a and 2b shows that the step height increases for larger fractional area exposure but the location of the step is approximately the same for constant areal dose.

We demonstrate below that our observed steps are due to trapped positive charge in the exposed area, which causes that area to deplete and invert at more negative voltages than the unexposed portion of the capacitors. Although steps have been previously reported in C-V data after electron beam irradiation, they were of different origin. In a study of Zaininger (1967), steps were observed after total capacitor exposure and were attributed to the enhancement of existing interface states concentrated around a single energy level in the band gap. When the degree of perfection of the interface was high, no step was observed, but the entire C-V curve was shifted to more negative voltages. In order to test whether our steps were due to such a surface state phenomenon, we performed whole area exposures of our capacitors. The results of our test are shown in Figure 3. The shift of the C-V characteristics with no step present indicates that our steps in the partial exposure experiments are not related to the enhancement of pre-existing interface states but are a result of the localized positive charge created in the exposed area.

We now develop a quantitative model, which relates the size of the observed steps to the areas of exposures, assuming that the created positive charge is fixed where created. A single capacitor subjected to partial area exposure can be modeled as two capacitors connected in parallel: one capacitor, C_e , with an area equal to that of the exposed area, and another, C_u , with an area equal to that of the un-

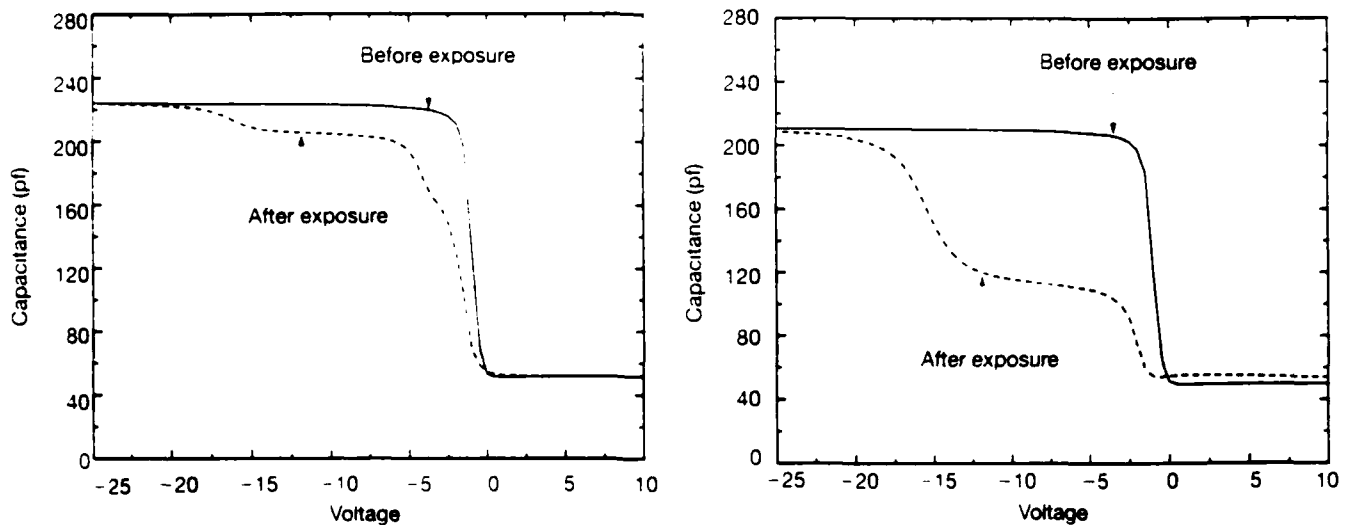


FIG. 2 C-V characteristics of MOS capacitors before and after e-beam partial area exposure. The exposure areas were (a) 10.4% and (b) 62.4% of the total area.

exposed area. Since the exposed area has a more negative threshold voltage, V_{ie} , because of the presence of the positive charge in the oxide, it undergoes depletion and inversion, while the unexposed area is still in accumulation. The unexposed area inverts at the unchanged threshold voltage, V_{iu} . Figure 4 shows schematically a C-V curve, which is a superposition of the individual C-V curves for the capacitors C_e and C_u . The additional step observed in the experiments is expected to occur at the threshold voltage of the exposed area. The height of the step should be equal to the difference between the accumulation capacitance, C_{ae} , and the inversion capacitance, C_{ie} , of the exposed area.

To determine the theoretical relationship between the step height, $C_{ae} - C_{ie}$, and the exposed area, A_e , we con-

sider the following relations as derived from basic MOS C-V theory (Pierret 1983). The accumulation capacitance of the exposed area is given by,

$$C_{ae} = \frac{\kappa_{ox} \epsilon_0 A_e}{d_{ox}} \quad (1)$$

where κ_{ox} is the dielectric constant of SiO_2 , ϵ_0 is the vacuum permittivity, and d_{ox} is the oxide thickness. Similarly, the semiconductor capacitance of the exposed area in inversion can be expressed as:

$$C_{ie} = \frac{\kappa_{Si} \epsilon_0 A_e}{W} \quad (2)$$

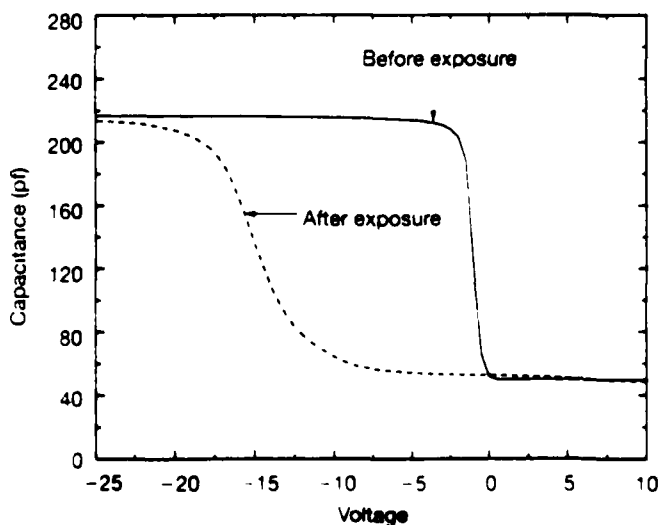


FIG. 3 C-V characteristics of a MOS capacitor before and 30 min after total e-beam area exposure.

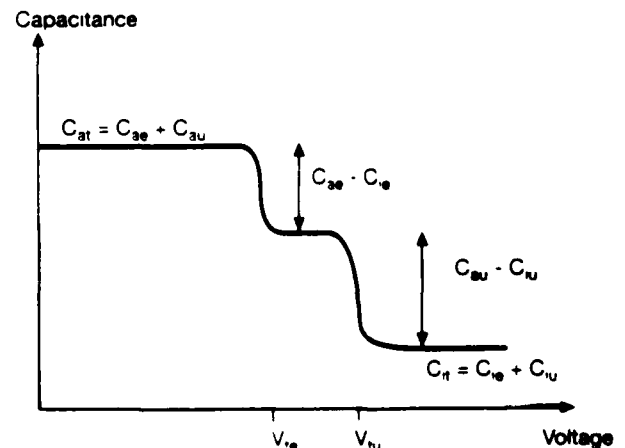


FIG. 4 Schematic diagram of the C-V curve resulting from partial area exposure. Key: C = capacitance, with subscripts a = accumulation, t = total, e = exposed, u = unexposed, and i = inversion. V_{ie} and V_{iu} are threshold voltages for exposed and unexposed portions, respectively.

where κ_s is the dielectric constant of silicon and W is the maximum depletion-layer width into the semiconductor. The capacitance of the exposed area in inversion can now be expressed as the series combination of the above two capacitances:

$$C_{ie} = \frac{C_{de} C_{se}}{C_{de} + C_{se}} \quad (3)$$

The total accumulation capacitance is:

$$C_{at} = \frac{\kappa_{ox} \epsilon_0 A_t}{d_{ox}} \quad (4)$$

where A_t is the total capacitor area. Thus, combining all of the above equations, we obtain a linear relationship between the step height and the exposed area:

$$\frac{C_{ae} - C_{ie}}{C_{at}} = K \frac{A_e}{A_t} \quad (5)$$

where

$$K = \left(\frac{1}{1 + \frac{\kappa_s d_{ox}}{\kappa_{ox} W}} \right)$$

Figure 5 compares the experimental data, obtained 30 min after exposure, and the theoretical predictions from Eq. (5). Shown is a plot of the normalized step height, $(C_{ae} - C_{ie})/C_{at}$, versus the normalized exposure area, A_e/A_t . The error bars in the experimental data represent uncertainty in the sharpness of the steps. For each individual capacitor with a certain exposed area, we also show the theoretical prediction from Eq. (5) for the step height, taking into account the precise value of the oxide thickness for the calculation of the constant, K . The straight line shown in the figure represents the relationship of Eq. (5) for a value of K corresponding to an average oxide thickness of 766 Å.

Figure 5 gives excellent agreement between our experimental results and the above theoretical model, which assumes that the positive charge remains fixed in the area where it is created. The step height in the C-V data may thus be used as a measure for the areal extent of positive charges in insulators. *In particular, the out-diffusion of positive oxide charges may be monitored as a change of the step height in time.*

To gain further insight into the time-dependence of the positive oxide charge, we performed an experiment to determine its position and magnitude a significant time after its creation. We analyzed the C-V data 30 min, 2 weeks, and 4 weeks after partial area exposure of an MOS capacitor, where the exposed area was 41.6% of the total area. The chip containing this capacitor was stored at room temperature in a dark shielded box during the 4-week time period. Figure 6 shows the before- and after-exposure curves. We see that the step observed 30 min after exposure has moved approximately 0.8 volts in the positive direction after 2 weeks, but only 0.2 volts more after an additional 2 weeks. Note that the step height did not

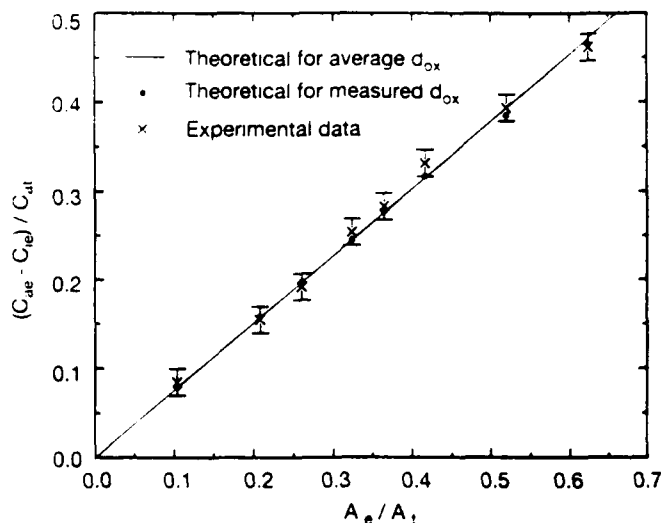


FIG. 5 Plot of theoretical and experimental results relating the exposure area to the resulting step in the C-V characteristics. The crosses with error bars indicate experimental data. The filled circles represent the predictions of our theoretical model using the actual oxide thickness for each case. The solid line shows the linear relationship of Eq. (5) for the average oxide thickness of $d_{ox} = 766$ Å.

change over the 4-week period, implying that *no measurable diffusion of charge has taken place.*

Even though the charge did not spread out, its magnitude decreased over time. The amount of charge dissipation can be determined from simple MOS theory. The shift in the threshold voltage of the exposed area, ΔV_{te} , is related to a change in the oxide charge, ΔQ_{ox} , by:

$$\Delta V_{te} = \frac{-\Delta Q_{ox}}{C_{ae}} \quad (6)$$

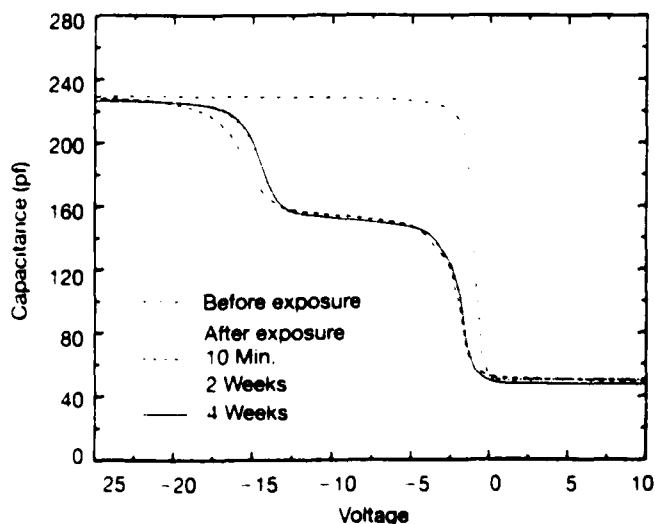


FIG. 6 C-V characteristics of a MOS capacitor before exposure and 30 min, 2 weeks, and 4 weeks after 41.6% of the total area was exposed with an electron beam.

The shift in the threshold voltage of -14.5 volts 30 min after the exposure corresponds to an induced charge density of 6.8×10^{-7} C/cm². After 2 weeks, the charge is reduced by only about 6% to 6.4×10^{-7} C/cm². After an additional 2 weeks, the charge is further reduced a mere 1.5%. Positive fixed charges in the oxide can be considered nearly "permanent" in the absence of any annealing treatments. In fact, Zaininger (1967) observed that positive charge can endure for several months in capacitors wholly exposed to electron radiation.

One might theorize a mechanism by which induced positive charge is not fixed in the oxide but, in fact, leaks out over a small distance onto the ungrounded metal (i.e., electrons into the oxide), causing a bending of the band structure for the two materials and resulting in a cessation of charge transfer after some time. Such a mechanism might not be detectable by our technique, as our lateral resolution is obviously limited and the oxide is very thin. However, if this were the case, the metal probe used in the C-V measurement would quickly drain the charge for the metal and allow a continued leakage of charge to the C-V measurement system. Presumably, such a leakage would be expected to act rapidly since in an SEM, sample charging is apparently controlled on time scales of the order of screen refresh rates, that is, a matter of a few seconds or less.

Conclusion

We have shown that partial area exposure of MOS capacitors using electron beam irradiation results in the creation of an area of *immobile* positive charge within the SiO₂ layer. This charge causes depletion and inversion of the exposed area of the capacitor at more negative voltages than the unexposed area, which resulted in a step in the C-V characteristics. A systematic study of exposed area as a function of step height indicated that step height increased with increasing exposure area. After a theoretical model had been proposed, experimental results were compared with those obtained from this model. The good agreement we obtained demonstrated a direct correlation between the exposed area and the step height in the C-V data. We exploited this phenomenon to show that the positive oxide charge does not diffuse over long time periods. This demonstrates that in electron microscopy, the common technique of coating nonconductive substrates with a metal overlayer in the case of positive charging is accomplished not by the draining of excess charge away from the

imaging site but rather by the electrical shielding of the site by a ground plate.

Acknowledgments

The authors thank Mr. Robert J. Minniti, Jr., and Mr. Malay Kundu for assistance with measurements and data acquisition, and Mr. Greg Bazán for electron beam lithography exposures. We are grateful to Prof. Stephen J. Krause and Prof. David C. Joy for helpful discussions.

References

- Balasinski A, Jakubowski A, Swit A: The influence of electron beam energy on defect density in MOS device quality oxides. *Vacuum* 38, 1041-1043 (1988)
- Goldstein JI, Newbury DE, Echlin P, Joy DC, Fiori C, Lifshin E: *Scanning Electron Microscopy and X-Ray Microanalysis*. Plenum Press, New York (1981) 461-494
- Henderson RC: Device fabrication by electron beam lithography. In *Electron Beam Technology in Microelectronic Fabrication*. (Ed. Brewer GR) Academic Press, New York (1980) 226-227
- Johnson WC: Mechanisms of charge build-up in MOS insulators. *IEEE Trans Nucl Sci* NS-22, 2144-2150 (1975)
- Keery WJ, Leedy KO, Galloway KF: Electron beam effects on microelectronic devices. *Scanning Electron Microscopy, Proc Workshop on Microelectronic Device Fabrication with the SEM*, IIT Research Institute, Chicago (1976) 507-513
- Krause SJ, Mohr J, Bernstein G, Ferry DK, Joy DC: A new method for measuring the thickness of thin film insulators. *Proc Microbeam Analysis Soc* 14, 459-462 (1989)
- McLean FB, Boesch HE Jr, Oldham TR: Electron-hole generation, transport, and trapping in SiO₂. In *Ionizing Radiation Effects in MOS Devices and Circuits* (Eds. Ma TP, Dressendorfer PV) John Wiley & Sons, New York (1990) 87-92
- Nicollian EH, Brews JR: *MOS (Metal Oxide Semiconductor) Physics and Technology*. John Wiley & Sons, New York (1982) 549-553
- Pierret RF: Field effect devices. *Modular Series on Solid State Devices* 4, Addison-Wesley, (1983) 43-80
- Sah CT, Sun JYC, Tzou JYT: Effects of keV electron radiation on the avalanche-electron generation rates of three donors on oxidized silicon. *J Appl Phys* 54, 1563-1568 (1983)
- Shimaya M, Shiono N, Nakajima O, Hashimoto C, Sakakibara Y: Electron beam induced damage in poly-Si gate MOS structures and its effect on long-term stability. *J Electrochem Soc* 130, 945-950 (1983)
- Zaininger KH: Irradiation of MIS capacitors with high energy electrons. *IEEE Trans Nucl Sci* NS-13, 237-247 (1966)
- Zaininger KH, Holmes-Siedle AG: A survey of radiation effects in metal-insulator-semiconductor devices. *RCA Review* June 208-240 (1967)

Swelling and surface forces-induced instabilities in nanoscopic polymeric structures

Davide A. Hill^(*)

Department of Chemical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

Xiaokang Huang, Greg Bazán, and Gary H. Bernstein

Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

(Received 1 May 1992; accepted for publication 12 July 1992)

For closely spaced, nanolithographically defined lines, a thin wall of resist remains to act as the metal line spacer. When exposed to a developer, and then rinsing solution, closely spaced resist walls may become unstable as a result of two effects: (1) internal stresses due to swelling, and (2) lateral surface forces between adjacent walls. In this article we perform a linear stability analysis of a thin polymer wall under the simultaneous action of internal stresses and lateral surface forces. We calculate a stability boundary, and show that internal stresses are necessary for the formation of deformation patterns of finite wavelength. We find that, for slightly subcritical swelling stresses a small lateral force can induce buckling, while, for slightly subcritical surface tractions large internal stresses are necessary to induce instability. The theoretical predictions are in good agreement with experimental data on poly(methyl-methacrylate) walls produced by electron beam lithography.

INTRODUCTION

In addition to ultrahigh resolution, an important aspect of nanofabrication for nanoelectronics is the achievable spacing of metal features. Extremely closely spaced features have been utilized in such devices as lateral tunneling transistors,¹ split gates,² and optical gratings.³ In applications such as these, metal features are required with spaces as small as a few hundred Å. The fabrication of very narrow gaps over appreciable distances (a few μm) requires very good control of lithographic processing parameters.

Most pattern transfer in nanolithography for quantum devices is effected by the lift-off process in which a thin metal layer is deposited by perpendicular, thermal evaporation onto a resist [usually poly (methyl methacrylate) (PMMA)] such that metal falls either on the surface of the undeveloped (positive) resist or onto the surface of the substrate where exposed and developed resist has been removed. Subsequent removal of unexposed resist also removes all metal except that on the surface of the substrate. When metal patterns are to be placed very closely together, the space is formed by the shadowing of the deposited metal by a thin "wall" of resist. We report here on the fundamental limitations of solvent/PMMA interactions which lead to failure of the thin features necessary to perform lift-off of very closely spaced features.

For thin resists at moderate energies, undercutting by beam scattering within the resist can approximately be described by:⁴

$$b = 625(Z/E_0)t^{3/2}(\rho/A)^{1/2}, \quad (1)$$

where b is the increase of the diameter of the beam at the bottom surface of the resist, Z , A , ρ , and t are the atomic

number, atomic weight, density, and thickness of the resist, respectively, and E_0 is the beam energy in keV. For PMMA, $Z=3.6$, $A=6.7$ g/mol, and $\rho=1.2$ g/cm³.⁵ For 45 keV electrons and a PMMA thickness of 130 nm, undercutting is about 5 nm at each sidewall of our pattern, which aids in the lift-off process.

As discussed above, the unexposed resist spacer is in the form of a thin wall between the developed lines. When lines are placed too closely together and high resolution is achieved, the resist walls can either become wavy, fall over, or even lose adhesion to the surface, in which case they fail completely. In this regard, internal stresses due to swelling and lateral surface forces between adjacent structures can play dominant roles. Although PMMA suffers little swelling, which accounts partially for its ultrahigh resolution properties, some absorption of developer is necessary to aid the development process.⁶ Swelling and solubilization of polymeric resists have been extensively investigated.^{7,8} In marginally poor solvents, solvation forces favoring mixing are insufficient to promote dissolution of the polymer. Appreciable swelling can occur, however, leading to weaker mechanical properties, and to the buildup of internal stresses in constrained structures. Amorphous polymers of high molecular weight may typically experience between 0.1% and 30% swelling, without solubilization, depending on solvent quality and temperature.⁸ With mixtures of solvents, the swelling behavior can be of considerable complexity, since preferential partitioning of one or more components of the mixture in the polymer can occur. This would be the case for most resist developer solutions, which usually comprise more than a single component. Notwithstanding their valuable insight, thermodynamic models have enjoyed only limited success in the quantitative prediction of solubility properties of multicomponent polymer-solvent systems, and such an attempt will not be made here. Nevertheless, as a discriminator of the quality

^{*} Author to whom correspondence should be addressed.

of the model, our predictions will be required to compare favorably with typical swelling data.

Recent experiments have considerably clarified the mechanism of interaction between polymeric surfaces. Studies of this nature have recently become possible thanks to the perfecting of the surface force apparatus by Israelachvili and Tabor.⁹ In this system, the separation between two mutually orthogonal, cylindrical mica surfaces onto which polymer is adsorbed can be measured with resolution better than 5 Å. The separation measurements rely on interferometry, while knowledge of the flexural rigidity of the instrument enables calculation of the force between the surfaces. Normalization of the force by 2π times the mean radius of curvature of the mica sheets directly yields the surface energy of interaction.¹⁰ Experiments show that in good solvents, polymer surfaces always repel each other due to the tendency of the chains to mix most favorably with solvent molecules. Conversely, in poor solvents chain-chain interactions are more favorable, leading to strong long-range attraction. Experimentally, the attractive forces between polymer layers become significant within a distance of about twice the radius of gyration of the chains, which can extend to several hundred Å for high molecular weight polymers.¹⁰ Common post-development rinsing solutions are, by design, of poor thermodynamic quality. Therefore, in addition to swelling, attractive forces are expected to play a vital role in the failure of polymeric nanostructures.

EXPERIMENT

Our electron beam lithography (EBL) system consists of an Amray 1400 scanning electron microscope (SEM) with a minimum spot size of 35 Å controlled by a personal computer interfaced through a Pragmatic Systems 2201A 16-bit arbitrary wave form generator. We used PMMA with molecular weight 950 000 amu baked at 170 °C for 4 h for all exposures. Resist and developer parameters were very tightly controlled. Resist thickness was 105–180 nm as measured with an Alpha-Step 200 surface profiler. The developer used in all cases was a solution of MIBK:IPA (1:3) with the addition of 1% (MEK) by volume.¹¹ It has been shown that contrast results are extremely sensitive to variations in developer temperature due to the high activation energy of the PMMA in the developer solvents.¹¹ Using a temperature controller, our developer temperature was maintained at 25 ± 0.5 °C. Develop time for all samples was 15 s followed by a 15 s rinse in methanol. After development, all samples were dried in air at 90 °C for 5 min.

All exposures were performed at a beam energy of 45 keV at a spot size of 5 nm and 120 μm scan field. Substrates used in all cases were 200 nm SiO_2 on thick silicon wafers. Doses were very critical for making high-density patterns because of proximity effects during *e*-beam lithography. They were not only related to the line pitch, but also to the size of the patterns. For grating pitches ranging from 70 to 150 nm, the doses varied from 1.5 to 2.5×10^{-4} C/cm. For the case of only two parallel lines, doses ranged from 2.5 to 3.5×10^{-4} C/cm for pitches from 80 to 120 nm.

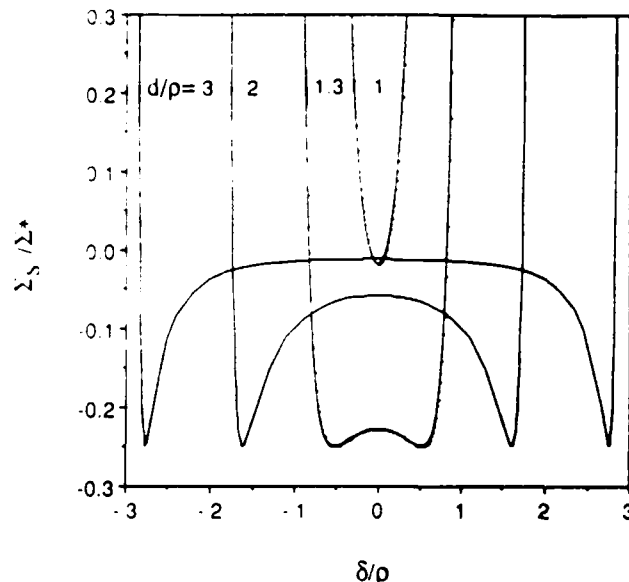


FIG. 1. Schematic representation of the potential energy of interaction of a polymer surface confined between and equidistant from two other surfaces. As the dimensionless separation d/ρ increases, the curves broaden and the equilibrium position around $\delta/\rho=0$ becomes indifferent to small lateral displacements. At very small separations the potential becomes purely repulsive. At intermediate separations the erect position may either be stable or unstable depending on the rigidity of the wall.

All metal evaporations were gold from an electron beam evaporator with thicknesses of 30 ± 1 nm. Lift-off was performed by soaking the samples in acetone for about 5 min, followed by the application of acetone using a hypodermic syringe, followed by a 15 s methanol rinse, and final blow drying in nitrogen.

THEORY

The interaction energy between two opposing surfaces, Σ , can be obtained by linear superposition of long-range, attractive, and short-range, repulsive interactions. Generally, the potentials are individually modeled as a power-law of the reciprocal of the separation distance, r , giving (within an arbitrary additive constant):^{10,12}

$$\Sigma = \Sigma^* \left[\left(\frac{\rho}{r} \right)^a - \left(\frac{\rho}{r} \right)^b \right], \quad (2)$$

where Σ^* and ρ characterize the strength and width of the potential well, respectively, and a and b ($a > b$) prescribe the rate of growth and decay of repulsive and attractive interactions, respectively. For a polymer wall confined between, and equidistant from two parallel surfaces with separation d , the following approximation is appropriate:

$$\Sigma_s = \Sigma^* \left\{ \left[\left| \frac{\rho^2}{(d^2 - \delta^2)} \right| \right]^{a/2} - \left[\left| \frac{\rho^2}{(d^2 - \delta^2)} \right| \right]^{b/2} \right\} \quad (\text{for } -d < \delta < d), \quad (3)$$

where Σ_s is the surface energy (energy/area), and δ any virtual lateral displacement away from the planar configuration. Note that δ , and hence Σ_s , will in general vary from point to point along the surface of the wall. Figure 1 schematically illustrates the shape of Σ_s for different values of the dimensionless separation, d/ρ . It is seen that for

large values of d/ρ the erect position is indifferent, while it becomes progressively more unstable as the surface-to-surface separation decreases. (Note that we are ignoring for the moment the elasticity of the wall, which, by counteracting deformation, would stabilize the system.) Interestingly, as the separation further decreases repulsion becomes dominant and the equilibrium reverts to unconditionally stable. Experimentally,¹⁰ the transition should occur around $d \approx R/2$, where R is the radius of gyration of the polymer chains. Apparently, this situation could be exploited to greatly improve resolution in nanolithography. However, it may be difficult in practice to take advantage of the phenomenon, since typically, $R/2$ is on the order of 5–10 nm. Additionally, interpenetration of the chains will, in fact, impede complete separation of the walls after solvent evaporation.

Differentiation of the interfacial energy [Eq. (3)] with respect to δ provides the interfacial stress acting on the lateral surface of the wall, σ :

$$\sigma = -\frac{d\Sigma_s}{d\delta} \quad (4)$$

In the limit of small displacements, the detailed structure of the potential is irrelevant, and σ can be linearized around the origin ($\delta=0$) to give

$$\sigma = \Theta \delta, \quad (5)$$

where

$$\Theta \equiv -\left(\frac{d^2\Sigma_s}{d\delta^2}\right)_{\delta=0} \quad (6)$$

Henceforth, Θ will be referred to as the "surface stress coefficient." Equation (6) can be used to calculate Θ from surface force measurements, for any given polymer-solvent system. Note that both the magnitude and sign of Θ will generally depend on the dimensionless separation d/ρ . Therefore, in the following we shall attempt to estimate only the order of magnitude of Θ .

Dimensional analysis of Eq. (6) and experiments¹⁰ suggest the following scaling:

$$\Theta \approx \Delta\Sigma_s/(R)^2, \quad (7)$$

where $\Delta\Sigma_s$ is the depth of the attractive potential well, and R the end-to-end distance of the polymer chains. As mentioned previously, surface-force-apparatus experiments directly yield Σ_s as the measured force normalized by 2π times the mean radius of curvature of the mica sheets.¹⁰ Unfortunately, experimental data of this nature are not available for PMMA. Polystyrene "brushes" in various solvents have been extensively investigated, however. Because of the osmotic nature of the interactions, the detailed chemical structure of the polymer is believed to be only of secondary importance.¹⁰ Therefore, data on polystyrene should prove useful for obtaining estimates of the general magnitude of $\Delta\Sigma_s$. Figure 1 of Ref. 10 gives plots of normalized force versus separation for polystyrene brushes in an aliphatic solvent below the θ temperature. The data show a pronounced attractive well with a depth of about

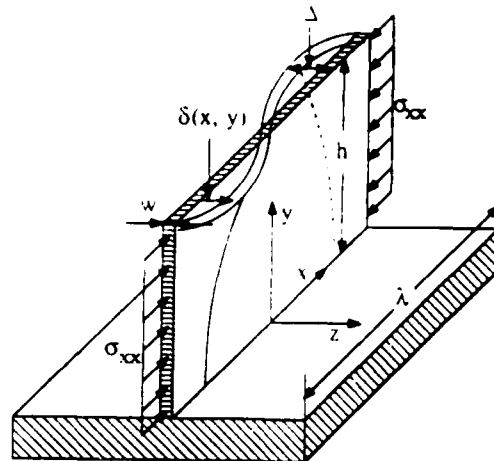


FIG. 2. Diagram of tall, thin wall demonstrating geometrical parameters used in the analysis. The wall has been hypothetically sectioned at the nodal points of the distortion.

$\Delta\Sigma_s \approx 160$ ($\mu\text{N/m}$) (after proper normalization by 2π). Roughly, we may assume $R \approx 10\text{--}30$ (nm) in Eq. (7), thus placing Θ in the range: $10^{11} < \Theta < 10^{12}$ (N/m^3). We show later that these values are sufficient to cause instabilities in arrays of closely spaced, thin polymer walls. In the following, we shall focus on the case $\Theta > 0$ (signifying an unstable equilibrium), and make use of Eq. (5) to describe the lateral surface traction on the wall.

We consider now the stability of an infinitely long wall of width w and height h under the simultaneous action of lateral forces and internal stresses due to swelling. The system of interest is shown in Fig. 2, which also gives the characteristic geometrical parameters. Mathematically, this problem is conveniently formulated within the framework of linear plate theory, leading to the following homogeneous boundary value problem for the lateral displacement, δ :¹³

$$\frac{\partial^4 \delta}{\partial x^4} + 2 \frac{\partial^4 \delta}{\partial x^2 \partial y^2} + \frac{\partial^4 \delta}{\partial y^4} = -\Lambda \frac{\partial^2 \delta}{\partial x^2} + \Omega \delta, \quad (8)$$

$$\delta = 0 \quad \text{at} \quad y = 0 \quad (\text{no displacement at the bottom}), \quad (9)$$

$$\frac{\partial \delta}{\partial y} = 0 \quad \text{at} \quad y = 0 \quad (\text{built-in bottom end}), \quad (10)$$

$$\frac{\partial^2 \delta}{\partial y^2} + \nu \frac{\partial^2 \delta}{\partial x^2} = 0 \quad \text{at} \quad y = 1 \quad (\text{torque-free top}), \quad (11)$$

$$\frac{\partial^3 \delta}{\partial y^3} + (2 - \nu) \frac{\partial^3 \delta}{\partial x^2 \partial y} = 0 \quad \text{at} \quad y = 1 \quad (\text{shear force-free top}) \quad (12)$$

where the dimensionless variables and parameters are defined as

$$x = X/h; \quad y = Y/h, \quad (13)$$

$$\Lambda = \frac{\sigma_{xx} 12(1 - \nu^2) h^2}{E w^2}, \quad (14)$$

$$\Omega = \frac{\Theta 12(1-\nu^2)h^4}{Ew^3} \quad (15)$$

In Eqs. (14) and (15), $\sigma_{xx} > 0$ is the internal compressive stress, Θ is defined in Eq. (5), ν is Poisson's modulus, E Young's modulus, and h and w the wall height and width, respectively.

For an infinitely long wall, the solution to Eq. (8) is of the form:¹³

$$\delta(x,y) = \cos(\xi x) \{A[\sinh(\alpha y) - \sin(\beta y)] + B[\cosh(\alpha y) - \cos(\beta y)]\}, \quad (16)$$

where A and B are integration constants, and

$$\alpha = \sqrt{\xi^2 + \sqrt{\Lambda\xi^2 + \Omega}} \quad (17)$$

$$\beta = \sqrt{-\xi^2 + \sqrt{\Lambda\xi^2 + \Omega}}.$$

Variable ξ is related to the wavelength of the prevailing distortion in the x direction, λ , as

$$\lambda = 2\pi h / \xi. \quad (18)$$

Imposition of the boundary conditions generates the following equation for the eigenvalues:

$$2ts + (s^2 + t^2) \cosh(\alpha) \cos(\beta) = \frac{1}{\alpha\beta} (\alpha^2 t^2 - \beta^2 s^2) \times \sinh(\alpha) \sin(\beta), \quad (19)$$

where

$$t = \beta^2 + \nu\xi^2$$

$$s = \alpha^2 - \nu\xi^2. \quad (20)$$

Note that in order for β in Eq. (17) to be real, we must have

$$0 < \epsilon < 1$$

where

$$\epsilon \equiv 2\xi^2 / (\Lambda + \sqrt{\Lambda^2 + 4\Omega}). \quad (21)$$

The use of ϵ , instead of ξ , provides a more efficient strategy for a systematic search of the eigenvalues.

Equation (19) implicitly defines a relation among Λ , Ω , and ϵ . In the search for the roots, the following limiting cases are important: (1) $\Lambda = 0$ and (2) $\Omega = 0$. These cases will be treated separately, since they provide clues for the solution of the general problem in which lateral forces and internal stresses act simultaneously.

Case (1): $\Omega = 0$

As is customary in stability theory, Eq. (19) (with $\Omega = 0$) is satisfied only by a discrete set of ϵ (or ξ) values, provided Λ is greater than or equal to a minimum critical value, Λ^* . Physically, Λ^* is the smallest (dimensionless) stress necessary to reach the marginal stability limit (indifferent equilibrium), and can be calculated by solving Eq. (19) numerically. We find: $\Lambda^* = 10.40$, and $\xi^* = 1.917$. Combining these results with Eqs. (14) and (18), the smallest critical stress, σ_{xx}^* , and the associated wavelength, λ , are calculated to be

$$\sigma_{xx}^* = 26Ew^2 / [30(1-\nu^2)h^2], \quad (22)$$

$$\lambda = h2\pi / 1.917. \quad (23)$$

For swelling induced stresses, σ_{xx}^* can in turn be obtained from

$$\sigma_{xx}^* = E\epsilon_{sw}, \quad (24)$$

where ϵ_{sw} is the hypothetical swelling strain that the wall would undergo if allowed to expand freely. Equations (22) and (24) enable calculation of the critical wall aspect ratio for stability, $(h/w)^*$, as a function of the swelling strain:

$$\left(\frac{h}{w}\right)^* = \sqrt{26/30(1-\nu^2)\epsilon_{sw}}. \quad (25)$$

Equation (25) is convenient for comparison with experiments, since it does not contain the unknown modulus E (which presumably changes upon swelling). As expected, a short wall can sustain a higher degree of swelling before buckling. Additionally, for a constant ϵ_{sw} , Eq. (25) predicts a direct proportionality between critical wall height and width. As shown later on, our experimental findings confirm this prediction.

Case (2): $\Lambda = 0$

Treatment of this case is similar to the previous one except that, here the smallest critical lateral "force," Ω^* , corresponds to $\xi^* = 0$. This implies an infinite wavelength, λ [Eq. (18)], which means that the wall does not buckle, but simply bends over sideways. Solving Eq. (19) numerically, we find: $\Omega^* \approx 12.363$, and $\xi^* = 0$. Making use of Eq. (15), the critical surface-stress coefficient, Θ^* , is obtained as

$$\Theta^* = \frac{1.03Ew^3}{(1-\nu^2)h^4}. \quad (26)$$

If estimates of E and ν were available, Eq. (26) could be used to determine Θ^* experimentally.

The search for the eigenvalues in the combined case (i.e., $\Lambda \neq 0$ and $\Omega \neq 0$) is now restricted to the domain ($0 < \Lambda < \Lambda^*$, $0 < \Omega < \Omega^*$, $0 < \epsilon < 1$). Delineation of the stability boundary implies the search of the minimum values $\tilde{\Lambda}$ and $\tilde{\Omega}$ that simultaneously satisfy Eq. (19). A convenient numerical strategy is to discretize the interval $0 < \tilde{\Lambda} < \Lambda^*$, and then find the corresponding minimum values of $\tilde{\Omega}$ (or vice versa). The results of these calculations are given in Table I, while Fig. 3 shows the stability boundary. The wavelength of the distortion, λ , as a function of Ω is shown in Fig. 4. Note that λ is relatively insensitive to lateral forces up to about 80% of Ω^* . It is interesting to note in Fig. 3 that, for $\Lambda < 5$, a slightly subcritical lateral stress, say $\Omega = 0.95 \Omega^*$, does not cause instability. On the other hand, for a slightly subcritical value of Λ , a small lateral force can be catastrophic.

RESULTS AND DISCUSSION

Plotted in Fig. 5 are the experimental critical heights for buckling versus width, for PMMA walls fabricated by electron beam lithography. A line through the origin satisfactorily fits the data, indicating a direct proportionality

TABLE I. Minimum values of Λ and $\Omega = \Omega(\Lambda)$ for stability, and the associated normalized wavelength, λ/h (the calculations assume $\nu=0.50$).

Λ	Ω	λ/h
10.40	9.000	3.277
10.19	1.000	3.396
9.990	2.000	3.582
9.737	3.000	3.623
9.474	4.000	3.653
8.932	6.000	4.002
8.272	8.000	4.319
7.450	10.00	5.097
6.700	11.30	6.398
6.000	12.08	9.721
5.000	12.362	759.2
0.000	12.363	∞

between critical wall height and width. This finding is in agreement with Eq. (25), assuming a constant degree of swelling. The critical aspect ratio for buckling can be calculated from the slope of the line, and we find that $(h/w)^* = 4.75$. Calculation of ϵ_{sw} from Eq. (25) requires knowledge of Poisson's modulus, ν . Because the polymer is swollen, and the internal stresses are small, the assumption $\nu=0.50$ is appropriate, giving $\epsilon_{sw}=5.1\%$. This value is well within the expected range for glassy polymers.

Another easily accessible observable is the wavelength of the distortion after buckling, λ , [Eq. (23)]. Figure 6 shows an electron micrograph of the shadow of a wall formed from a 140-nm-thick PMMA layer. The light areas are gold films created by lift-off. The measured wavelength of the distortion is 460 nm, in excellent agreement with the value of 459 nm given by Eq. (23) with $h=140$ nm.

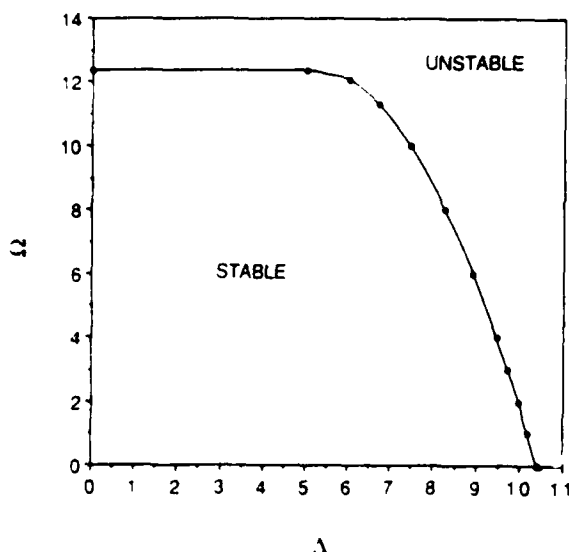


FIG. 3. Stability envelope for the case in which internal stresses (Λ) and lateral attractive forces (Ω) act simultaneously [Eqs. (14) and (15)]. For $\Lambda \leq 5$ the curve is practically horizontal, implying that for slightly subcritical values of Ω , large internal stresses are needed to cause instability. On the other hand, for slightly subcritical values of Λ a small lateral force is sufficient to induce buckling.

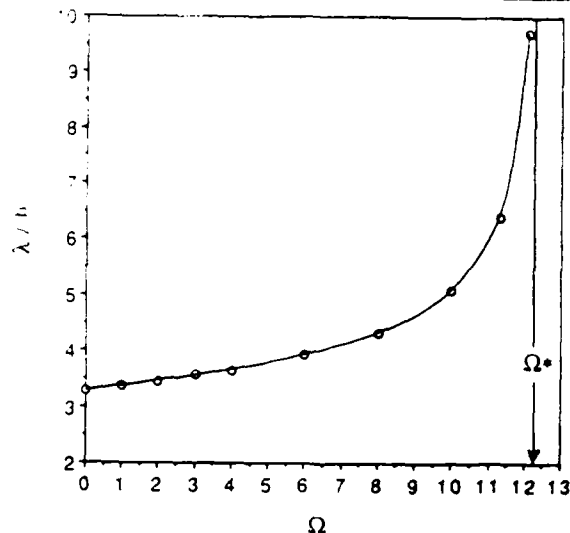


FIG. 4. Plot of the dimensionless wavelength of the distortion of the wall, λ/h , as a function of the dimensionless critical lateral stress, Ω .

The maximum amplitude of the lateral displacement of the wall Δ (see Fig. 2) can also be measured from the figure. The hypothesis that the deformation profile remains unaffected by solvent evaporation allows us to obtain an independent estimate of ϵ_{sw} from the observed value of Δ . The curvilinear contour length between nodal points at the top of the wall after buckling can be calculated from

$$L = \lambda(1 + \epsilon_{sw}) = \frac{\lambda}{2\pi} \int_0^{2\pi} \sqrt{1 + \left(\frac{2\pi\Delta}{\lambda}\right)^2 \sin^2(z)} dz. \quad (27)$$

For $(2\pi\Delta/\lambda)^2 \ll 1$ (subject to verification), the integration kernel in Eq. (27) can be expanded up to second order terms and integrated to give

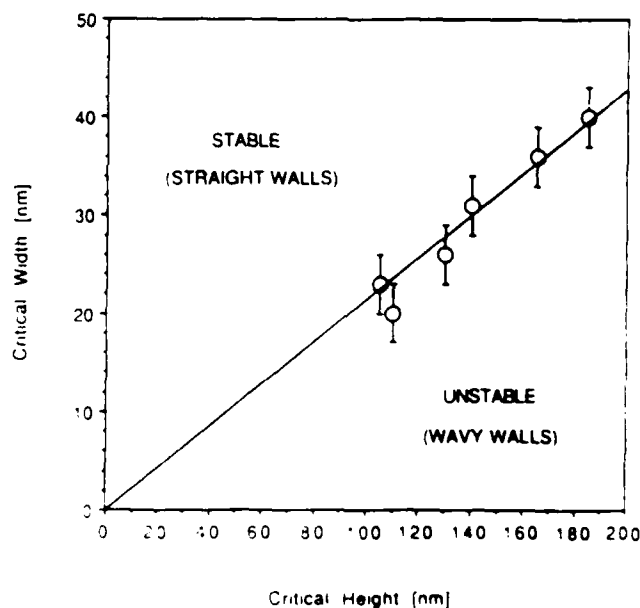


FIG. 5. Minimum width of PMMA walls achievable for a given starting resist thickness (wall height).

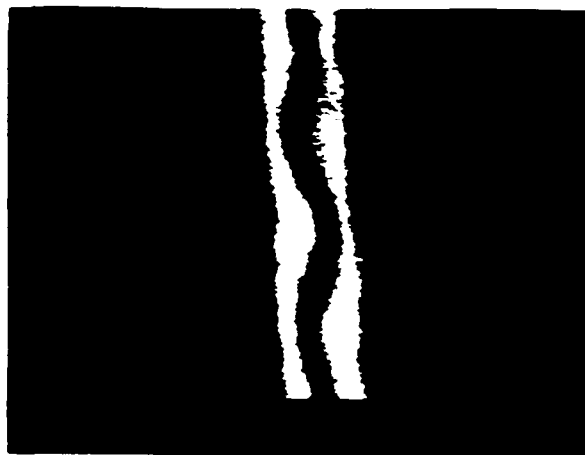


FIG. 6. Remaining metal after lift-off of a typical failed wall. Light areas are gold, and dark, wavy line is the shadow of failed PMMA wall. The scale marker is 100 nm.

$$\epsilon_{sw} \approx \left(\frac{\pi \Delta}{\lambda} \right)^2 \left[1 - \frac{3}{2} \left(\frac{\pi \Delta}{\lambda} \right)^2 \right]. \quad (28)$$

This equation can be used to calculate ϵ_{sw} , with the measured values of Δ and λ . From Fig. 6 we estimated $\Delta \approx 33$ nm, giving (with $\lambda = 460$ nm) $\epsilon_{sw} = 4.7\%$. The agreement with the previous estimate of 5.1% is satisfactory. These findings confirm that swelling is the dominant mechanism for instability in our experiments. It is remarkable that a 5% swelling can lead to the failure of a wall less than five times higher than its width.

In the above experiments the effects of lateral surface forces were presumably small, due to the relatively large separation between features. As pointed out at the end of the previous section, for slightly subcritical values of Λ , a small lateral force is sufficient to induce instability. In this regard, the magnitude of parameter Θ in Eq. (15) is expected to be of crucial importance. Previously obtained estimates of Θ ranged from 10^{11} to 10^{12} (N/m³). Therefore, we may verify now whether or not instabilities can be triggered for physically meaningful values of Θ .

The worst possible scenario envisions surface forces alone as sufficient to induce lateral collapse of the wall (i.e., $\Lambda = 0$). Equation (26) can then be used to calculate a

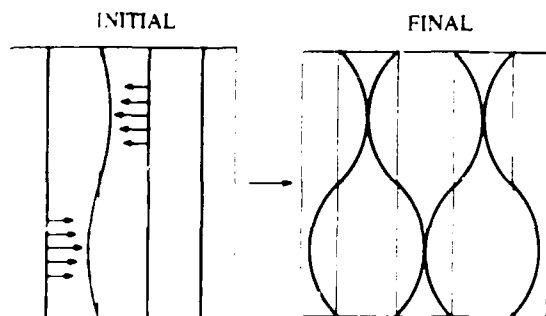


FIG. 7. Schematic illustration of lateral interactions between closely spaced features, immediately after swelling induced failure of one of the walls.

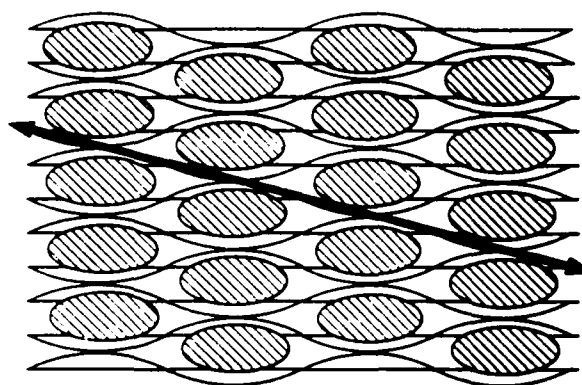


FIG. 8. Schematic of the final topology of the pattern in Fig. 7. The shaded ellipsoidal areas represent metal patches left after metal evaporation and lift-off.

critical wall height, h^* , for assumed values of E , ν , w , and Θ . Choosing $E = 10^2$ (MPa), $\nu = 0.5$, $w = 30$ (nm), and $\Theta = 10^{12}$ (N/m³), we estimate $h^* \approx 250$ (nm) which is in the range of interest for nanolithography.

The implications of these findings for fabrication of arrays of closely spaced walls can be substantial. Indeed, for a partially swollen polymer, surface forces can induce cross-correlation between deformation of adjacent structures, with important consequences on the final topology of the pattern. Figure 7 schematically illustrates this effect. It is seen that the swelling induced buckling of one of the walls can lead to the catastrophic failure of the entire array, producing a pattern of adjacent sinusoidal waves 180° out-of-phase with respect to each other. Metallization followed by lift-off would then cause the appearance of a pattern of metal islands, as shown in Fig. 8. Experimental data supporting these ideas are shown in Fig. 9. Therefore, attractive surface interactions can be extremely important in the fabrication of thin, closely spaced features. This issue



FIG. 9. Scanning electron micrograph showing a failed pattern of an array of closely spaced walls. The dark diagonal bands are the loci of the points where failed resist walls have come into contact (prior to metallization). The scale marker corresponds to 1 μ m. This figure should be compared with Fig. 8.

is especially crucial when dealing with marginally poor solvents, as is the case for most post-development, rinsing solutions.

As a final note on the modeling of wall stability, the following points must be considered. First, in the theoretical treatment we have ignored the effect of hydrodynamic forces which, at least in principle, could also be invoked as a source of instability (e.g., agitation during development). In fact, such effects are unimportant within the context of the stability problem, since upon cessation of flow a hypothetically deformed structure would always "spring back" to its most stable configuration (as dictated by the equilibrium criteria derived previously). Additionally, hydrodynamic effects cannot explain the correlation between deformation of adjacent structures observed in Fig. 9. Second, our simplified analysis assumes that the walls are: (1) strongly anchored at the bottom and (2) of constant width. The validity of the first assumption is easily inferred from the experiments, since floppy, weakly anchored walls would be incapable of retaining a deformed configuration of high strain energy such as the one shown in Fig. 6. In regards to the second assumption, we know that it cannot be entirely accurate, since undercutting inevitably leads to walls of decreasing thickness from top to bottom. A full analysis of this effect goes beyond the scope of the present work, but simple qualitative arguments in support of our approximation can be offered for the case of mildly tapered, strongly anchored walls in the swelling dominated regime (i.e., $\Omega \ll \Lambda$). Indeed, due to the anchoring boundary condition, the deformation at the bottom of the wall is small [see Eq. (16)]. Therefore, as a result of buckling, regions of the wall closer to the surface experience neither an appreciable release of strain energy nor a sizable decrease of surface free energy. In other words, the bottom portion of the wall contributes negligibly to the onset of instability, since its contribution to the overall energy balance is small. The relevant wall width, w , controlling the onset of the instability [i.e., the value to be used in Eqs. (14) and (22)] should then be closest in value to that at the top, free end. This argument is also consistent with the fact that, for $\Omega \ll \Lambda$, the wavelength of the distortion, λ [Eq. (23)], is independent of w .

SUMMARY AND CONCLUSIONS

Upon exposure to suitable solvents such as common developing and rinsing solutions, nanofabricated polymeric walls may become unstable as a result of two effects: (1) swelling and (2) lateral surface forces between adjacent structures. By performing a linear stability analysis we cal-

culated the stability envelope and wavelength of the distortion for the general case in which internal swelling stresses and lateral surface forces act simultaneously. We also showed experimental evidence that supports our predictions, thus demonstrating the existence and severity of such instabilities in nanolithography. In the limiting case of small surface forces, predictions and data agreed quantitatively. Order-of-magnitude calculations of the strength of the lateral interactions between polymer walls (based on experimental surface force data) confirmed that lateral attractive forces can substantially enhance the action of swelling stresses, leading eventually to failure. Practical implications of these instabilities in the fabrication of arrays of closely spaced polymer walls were discussed. Formulation of developers that maximize contrast, but minimize swelling and attractive interfacial forces, is thus relevant to attempts to achieve higher resolution in nanostructure fabrication.

ACKNOWLEDGMENTS

We would like to thank D. T. Leighton for helpful discussions and S. Subramaniam for help in sample preparation. This research was supported in part by IBM, AFOSR, ONR, and the University of Notre Dame. D. A. H. gratefully acknowledges partial support by the Donors of The Petroleum Research Fund, administered by the American Chemical Society.

- ¹ D. R. Allee, S. Y. Chou, J. S. Harris, Jr., and R. F. W. Pease, *J. Vac. Sci. Technol. B* 7, 2015 (1989).
- ² J. C. Wu, M. N. Wybourne, W. Yindeepol, A. Weisshaar, and S. M. Goodnick, *Appl. Phys. Lett.* 59, 102 (1991).
- ³ S. Y. Chou, Y. Liu, and P. B. Fischer, *IEDM* 91, 745 (1991).
- ⁴ J. I. Goldstein, in *Introduction to Analytical Electron Microscopy*, edited by J. J. Hren, J. I. Goldstein, and D. C. Joy (Plenum, New York, 1979), p. 101.
- ⁵ G. Owen, *Repts. Prog. Phys.* 48, 795 (1985).
- ⁶ A. Weill, in *The Physics and Fabrication of Microstructures and Microdevices*, edited by M. J. Kelly and C. Weisbuch (Springer, New York, 1986), p. 58.
- ⁷ W. M. Moreau, *Semiconductor Lithography* (Plenum, New York, 1988), chap. 10.
- ⁸ G. Park and K. Ueberreiter, in *Diffusion in Polymers*, edited by J. Crank and G. Park (Academic, New York, 1968), pp. 140-162, 219-257.
- ⁹ J. N. Israelachvili and D. Tabor, *Proc. R. Soc. London Ser. A* 331, 19 (1972).
- ¹⁰ S. S. Patel and M. Tirrell, *Ann. Rev. Phys. Chem.* 40, 597 (1989).
- ¹¹ G. H. Bernstein, D. A. Hill, and W. P. Liu, *J. Appl. Phys.* 71, 4066 (1992).
- ¹² H. J. Taunton, C. Toprakcioglu, L. J. Fetters, and J. Klein, *Macromolecules* 23, 571 (1990).
- ¹³ S. P. Timoshenko and J. M. Gere, *Theory of Elastic Stability*, 2nd ed. (McGraw-Hill, New York, 1961), p. 363.

FEATURES OF QUANTUM MAGNETOTRANSPORT AND ELECTROMIGRATION IN MESOSCOPIC SYSTEMS

S. Bandyopadhyay, S. Chaudhuri and B. Das
Department of Electrical Engineering
University of Notre Dame
Notre Dame, Indiana 46556

M. Cahay
Department of Electrical and Computer Engineering
University of Cincinnati
Cincinnati, Ohio 45221

(Received 4 August 1992)

We have calculated the two-dimensional spatial distributions of several transport variables such as the current density, carrier concentration, chemical potential, space-charge electric field, residual resistivity dipole potential and electromigration forces in disordered mesoscopic structures subjected to arbitrary magnetic fields. These fully quantum mechanical calculations shed new light on such magnetotransport phenomena as the integer quantum Hall effect, the formation of magnetic bound states, magnetic response of current vortices caused by quantum interference between scatterers, the nature of residual resistivity dipoles in phase coherent transport, the difference between majority- and minority-carrier mobilities in the quantum mechanical regime, the electron wind force and direct force of electromigration in a mesoscopic system, and a variety of other phenomena.

Introduction

The spatial distribution of transport variables (such as the current density, carrier concentration, chemical potential, residual resistivity dipole field, etc.) are important in understanding many quantum transport phenomena¹ in mesoscopic structures and also electromigration effects^{2,3}. In this paper, we present a microscopic calculation of such spatial distributions in a quasi one dimensional structure containing localized impurities and subjected to arbitrary magnetic fields. The calculations are fully quantum-mechanical and are based on the Schrödinger equation describing the system. Our results elucidate the nature of many different types of quantum magnetotransport phenomena (e.g. the integer quantum Hall effect, current vortex formation, etc.), and they also shed light on electromigration forces, particularly the wind force and the direct force on an ion that cause electromigration in solids. Electromigration is an especially important phenomenon in narrow one-dimensional mesoscopic samples, not only in its extreme effect of causing catastrophic device failure as it does in integrated circuits, but also in that it causes impurity motion which radically alters sample characteristics and leads to 1/f noise.

This paper is organized as follows. In the next section, we describe the theory for calculating spatial distributions of transport variables in mesoscopic structures. We then present results dealing with the two-dimensional spatial distribution of the current density in a disordered structure at various magnetic fields. The current distributions show a number of dramatic effects such as *current crowding* near an attractive impurity, *current detour* around a repulsive impurity, the formation of edge states with perfect transmissivity at the onset of the *integer quantum Hall effect*, circulating current patterns around an impurity characteristic of the formation of *magnetic bound states*, formation of current vortices (in the absence of any magnetic field) as a result of quantum interference between waves reflected from various impurities in a sample, the quenching of such vortices by a magnetic field, and so on. These results and others are accompanied by the corresponding modulations in the carrier concentration profile caused by current flow. From these profiles, one can clearly see the building up of an excess electron concentration around an attractive scatterer and a deficit around a repulsive one owing to electrostatic interaction of the electrons with the scatterer. Also, electrons accumulate around an impurity (irrespective of whether it is attractive or repulsive) when

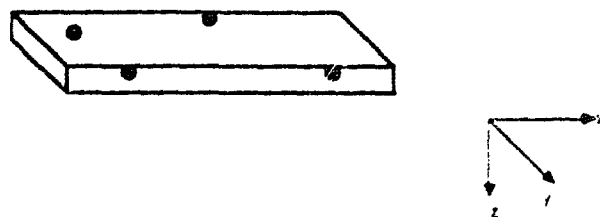
a magnetic bound state forms. These accumulations and depletions lead to inhomogeneities in the charge distribution which, in turn, lead to inhomogeneities in the electric field within the sample. Such inhomogeneities have a profound effect on the driving forces of electromigration acting on an impurity located inside the sample.

In addition to the spatial distributions of the current density and carrier concentration modulation, we also show the spatial distribution of the chemical potential inside a sample. In the absence of any magnetic field, the chemical potential profiles show very different features depending on whether the scatterers are attractive or repulsive. When a magnetic field is applied, the chemical potential profiles for both attractive and repulsive scatterers change dramatically. At high enough magnetic fields, the edges of the structure (along the direction of current flow) become strikingly smooth equipotential surfaces. This happens even when the potential at the center of the structure is exhibiting chaotic behavior. As a result, the longitudinal resistance measured by voltage probes attached to the edges of the structure will always read precisely zero (in a four-probe measurement). This is a clear demonstration of the quantum Hall effect. Also, we find that the chemical potential difference between the two edges of the structure (i.e. the Hall voltage drop) is exactly equal to the chemical potential drop between the two contacts of the structure when a single spin degenerate subband is occupied. Since the quantum mechanical transmission probability through the structure is exactly unity when this happens, it means that the Hall resistance is precisely quantized to $h/2e^2$ which is an unequivocal demonstration of the integer quantum Hall effect. It should be noted that these results pertaining to the integer quantum Hall effect are found from a rigorous microscopic calculation and *do not invoke any phenomenological construct*.

From the electrostatic potential profile, we have calculated the electric field distribution in the structure. The field shows significant inhomogeneities which have a profound influence on electromigration. Electromigration is caused by forces acting on charged impurities (ions) within the structure which tend to cause ion motion. Such motion can drastically alter the behavior of the structure in the phase coherent regime, lead to 1/f noise, and, in the extreme case, physically disrupt the structure leading to total device failure. The two main forces of electromigration are the "wind force" caused by collisions with electrons flowing past the ion and the "direct force" due to the electric field acting on the ion. We have calculated the spatial distribution of the so-called residual resistivity dipole potential which has a profound effect on the "wind force". In addition, the directions of both the "wind force" and the "direct force" at various ion locations and at varying magnetic field have also been calculated. To our knowledge, these are the first calculations of the effects of a magnetic field on electromigration forces in mesoscopic samples in the quantum coherence regime.

Theory

We consider a quasi one-dimensional structure as shown in Figure 1. The thickness is small enough that only a single



• Fig. 1: A quasi one-dimensional structure containing a random distribution of elastic scatterers. The structure is thin enough in the z -direction to allow the occupation of only one transverse subband in that direction. The magnetic field is in the z -direction.

transverse subband can be occupied in the z -direction (i.e. the confinement is complete in the z -direction). However, multiple transverse subbands may be occupied in the y -direction.

The Schrödinger equation describing this system under a z -directed magnetic field is

$$\frac{(\vec{p} - e\vec{A})^2}{2m^*} \psi + V\psi = E\psi, \quad (1)$$

where \vec{p} is the momentum operator, m^* is an isotropic effective mass, \vec{A} is the magnetic vector potential, $\psi(x, y)$ is the two-dimensional wavefunction, and the potential energy V consists of two terms

$$V \equiv V(x, y) = V_c(y) + V_{imp}(x, y). \quad (2)$$

The first term gives the effect of the confining potential in the y -direction and the second term is due to impurities. We assume hardwall boundary conditions in the y -direction and delta potentials for the impurities so that we have

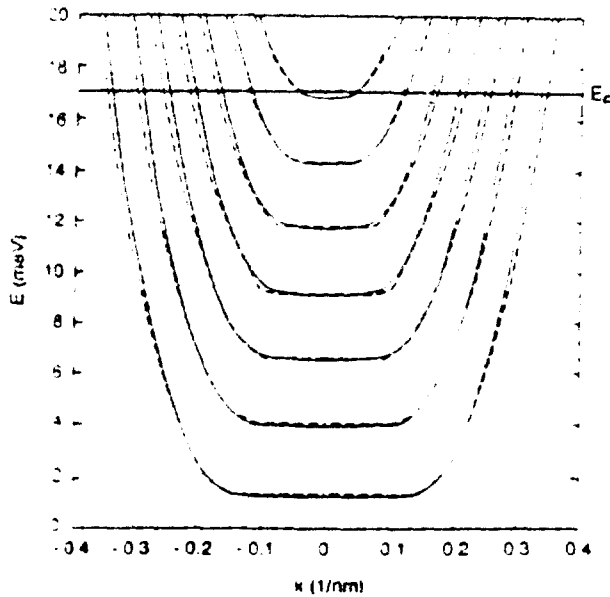
$$V_{imp}(x, y) = \gamma \sum_{i=1}^s \delta(x - x_i) \delta(y - y_i), \quad (3)$$

where s is the total number of impurities in the structure and the summation extends over all the impurities. The position of the i th impurity is denoted by (x_i, y_i) . In all our calculations, we neglect effects associated with the spin of the electron.

The confining potential in the y -direction and the z -directed magnetic field give rise to a set of hybrid magnetoelectric subbands in the y -direction. Each such subband corresponds to a mode or a transport channel. Calling $\phi_m(y)$ the y -component of the wavefunction in the m th magnetoelectric subband, we can write the total wavefunction $\psi_{E_F, m}(x, y)$ for an electron injected from the left contact with the Fermi energy E_F into the m th magnetoelectric subband as

$$\psi_{E_F, m}(x, y) = \sum_{p=1}^M \frac{A_{mp}(x) \exp(ik_p^F x) \phi_p(y)}{\sqrt{|v_p^F|} L} + \sum_{p=1}^M \frac{B_{mp}(x) \exp(ik_{-p}^F x) \phi_{-p}(y)}{\sqrt{|v_p^F|} L}, \quad (4)$$

where L is the length of the structure, and k_p^F and v_p^F are the x -directed wavevector and x -directed group velocity correspond-



• Fig. 2: The energy-dispersion relations for the hybrid magneto-electric subbands in a 2000 Å wide structure at a magnetic flux density of 1.5 Tesla. The solid curves are the result of an exact numerical calculation and the broken curves are those obtained from the application of Bohr-Sommerfeld quantization rule.

ing to the Fermi energy in the n th subband (see Fig. 2). These quantities are found from the dispersion relation of the hybrid magneto-electric subbands. The method for finding the dispersion relations, as well as the wavefunctions $\phi_n(y)$ in the n th magneto-electric subband, has been described in reference 4 and will not be repeated here. Finally, the subscript $-p$ in Equation (4) indicates that the quantity corresponds to a wavefunction with an oppositely directed velocity as compared to the one with a subscript p .

In Equation (4), the summation over p (subband index) extends over both occupied and unoccupied subbands (i.e. both propagating and evanescent modes). It is important to include enough evanescent modes in the summation so that the computed wavefunction does not depend on the number of terms (M) in the summation. Even though the evanescent modes do not carry current, they have very significant effects on all transport variables, and especially their spatial distributions¹.

When the wavefunction is written in the form of Equation (4), (i.e. each mode indexed by $\pm p$ is normalized by the square root of the corresponding group velocity), all propagating modes at the Fermi energy carry the same current. This is consistent with the well-known cancellation of the velocity and density of states in quasi one-dimensional structures which causes every mode to carry the same current. To find the total wavefunction given by Equation (4), we now only have to evaluate the coefficients $A_{mp}(x)$ and $B_{mp}(x)$. These are found from a scattering matrix method that has been described in the first two citations in reference 1. Once the wavefunction has been evaluated, we can calculate all quantities of interest such as the spatial distribution of the current, carrier concentration profiles, residual resistivity dipole potential, chemical potential, electric field, and

the forces of electromigration. The recipes for calculating these quantities from the wavefunction are described in the following paragraphs.

Spatial distribution of the current: In linear response transport and at low temperatures, only electrons with the Fermi energy carry a net current. Therefore, the two-dimensional wavefunction $\psi_m(x, y)$ of Fermi electrons in the m th magneto-electric subband allows us to calculate the contribution of that subband to the two-dimensional current density $\vec{J}_m(x, y)$.

$$\begin{aligned} \vec{J}_m(x, y) = & \frac{ie\hbar}{2m^*} \left\{ \psi_{E_F, m}(x, y) \nabla \psi_{E_F, m}^*(x, y) \right\} \\ & - \frac{ie\hbar}{2m^*} \left\{ \psi_{E_F, m}^*(x, y) \nabla \psi_{E_F, m}(x, y) \right\} \\ & + \frac{1}{m^*} (e\vec{A} |\psi_{E_F, m}|^2) \end{aligned} \quad (5)$$

The total current density is then found by vectorially adding the contributions from all occupied subbands (propagating modes).

$$\vec{J}_{total}(x, y) = \sum_m^{M'} \vec{J}_m(x, y) \quad (6)$$

The above expression provides the current density at every coordinate point (x, y) . Note that in calculating the total current density by summing over subbands, no weighting by the density of states is necessary in the summation since the wavefunctions have been normalized by the square root of velocity which takes care of the density of states factor. The above equation gives the two-dimensional spatial distribution of the current density (J_{total} as a function of coordinates (x, y)) for any arbitrary concentration and configuration of scatterers and for any magnetic field.

Carrier concentration change due to transport and its spatial distribution: To calculate the change in the carrier concentration caused by a current, we follow the method of Chu and Sorbello². At low enough temperatures, the deviation of the semiclassical electron distribution function from the equilibrium value in the m th subband (as caused by a small applied electric field \mathcal{E} driving the current) is given by

$$\Delta f_m(E) = -e\tau v(E, m) \mathcal{E} \delta(E - E_F) \quad (7)$$

where τ is a constant relaxation time describing a uniform background scattering due to all elastic collisions and $v(E, m)$ is the group velocity corresponding to electron energy E in the m th subband.

The change in the carrier concentration due to the current is then given by

$$\delta n_s(x, y) = \sum_m^{M'} \int dE N_{1d, m}(E) \Delta f_m(E) |\psi_{E, m}(x, y)|^2 \quad (8)$$

where $N_{1d, m}(E)$ is the one-dimensional density of states in the m th magneto-electric subband (the prescription for calculating this was provided in reference 4), M' is the total number of propagating modes or occupied subbands, and $\psi_{E, m}$ is the wavefunction of Fermi electrons in the m th magneto-electric subband. This wavefunction is not exactly the same as that given

by Equation (4); the only difference is that there should be no normalization of the various modes by the square-root of velocity in calculating $\Psi_{E,m}$. Therefore, if we use Equation (4) for the wavefunction, then $\delta n_i(x, y)$ should be written as

$$\delta n_i(x, y) = -\frac{1e\tau\mathcal{E}}{V} \sum_m |\Psi_{E,m} \Psi_{E,m}(x, y)|^2, \quad (9)$$

where ψ is now given by equation (4). The above equation gives the spatial distribution of the change in the carrier concentration as a consequence of transport.

Spatial distribution of the residual resistivity dipole potential: When an impurity is introduced in a solid, the resistance increases. This increase is associated with the formation of a "residual resistivity dipole" around the impurity which arises purely from the interaction of electrons with the impurity (scattering)¹. Landauer² has pointed out that the residual resistivity dipole potential is an important source of electric field inhomogeneities in a sample and affects the so called "wind force" of electromigration. The wind force is the force that acts on an impurity owing to collisions with electrons during current transport.

To calculate the spatial distribution of the residual resistivity dipole potential, we follow Chu and Sorbello². According to these authors, the electrostatic potential arising from δn , and the induced screening charge which attempts to locally neutralize δn , is the residual resistivity dipole potential. Within a Fermi-Thomas model, self-consistent screening gives this potential as

$$\delta U(x, y) = \frac{1}{\epsilon} \sum_m N_{i,m}^{-1} \delta n_{i,m}(x, y) - \frac{1}{\epsilon} \sum_m (E_F) |\Psi_{E,m}(x, y)|^2, \quad (10)$$

where $\delta n_{i,m}$ is the change in carrier concentration, from the m th magnetic subband.

From the above equation, we can readily obtain the spatial distribution of the residual resistivity dipole potential. The residual resistivity dipole field is the spatial gradient of this potential. In contrast to the claim of reference 3, this field does not have to be always directed along the direction of current flow, especially when quantum confinement (quantum size effect) and quantum interference (coherent effect) are important.

The "wind force" of electromigration: The wind force is the force that an ion or impurity in a solid experiences owing to an "electron wind" flowing past it during current transport. This electron wind imparts a force on the ion because the electrons collide with the ion and transfers momentum to the ion. The wind force causes electromigration (motion of ions or impurities) which has several effects in mesoscopic structures such as alteration of sample conductance, 1/f noise, and electrical failure caused by physical disruption of the sample at critical regions.

To calculate the wind force on an ion, we follow the work of Fiks and Huntington and Grone⁵ who invoke Newton's law. The wind force is given by

$$\vec{F}_{wind} = \left\langle \frac{d\vec{p}}{dt} \right\rangle_{coll}, \quad (11)$$

where the right hand side is the rate of loss of electron momentum due to collisions with the ion⁶. It can be shown⁷ from a rigorous semiclassical treatment based on the Boltzmann transport equation that the right hand side is $-\vec{J}_{total}/n\eta$ where n is the volume concentration and η is a semiclassical mobility associated with scattering by the ion. In the linear response regime, the above quantity is also equal to the negative of the spatial gradient of the chemical potential (quasi Fermi level). Therefore, the wind force on an ion located at coordinate (x_0, y_0) can be found directly from the chemical potential profile around the ion.

$$\vec{F}_{wind}(x_0, y_0) = -\vec{\nabla}\mu(x_0, y_0), \quad (12)$$

where μ is the chemical potential.

The above calculation of the wind force relies on the basic theory of reference 5 which does not distinguish between crystal momentum and real momentum in a solid. More sophisticated theories^{8,9} require knowledge of the exact shape of the ion potential and also the exact spatial distribution of the charge pile up (or charge deficiency in the case of repulsive scatterers) around ions. These sophisticated theories could actually benefit from our quantum mechanical calculations of the change in the charge density around the ions caused by current flow (described previously in this section). A fully quantum mechanical, self-consistent and rigorous calculation of the wind force, starting from the Schrödinger-Poisson equations, is reserved for a future publication.

The "direct force" of electromigration: The direct force on an ion in a solid subjected to an electric field is the electrostatic force given by

$$\vec{F}_{direct} = Ze\vec{E}, \quad (13)$$

where Z is the effective valency of the ion and \vec{E} is the local electric field at the ion site.

Bosvieux and Friedel⁸ claimed that the difference in the charge of an interstitial ion from the background will be completely screened by the electrons so that the local electric field and hence the direct force on such an ion will be exactly zero. Gupta and co-workers¹⁰ concluded in a similar vein that the direct force on a migrating ion in the saddle point position is exactly zero.

It was correctly pointed out by Das and Peierls¹¹ and also Landauer¹² that Bosvieux and Friedel could not be correct since their conclusion leads to inconsistencies. For instance, if an interstitial proton were completely screened and the local field is zero, then it should be also true of electrons and consequently a metal should have no electrical conductivity.

The correct equation to use for the direct force is Equation (13), but the electric field in that equation is not the space-averaged electric field as assumed by Huntington¹³ and by Lou et. al.¹⁴. It was stressed by Landauer that the field is affected by the "charge density modulation" around an impurity that occurs during current flow. Therefore, the field must be calculated self-consistently by solving the Poisson and Schrödinger equation. In the linear response regime, this field can be obtained without much difficulty as described below.

Spatial distribution of the chemical potential and electric field: Any net charge distribution $\rho(x, y)$ gives rise to an electrostatic potential $V(x, y)$ and an electric field $\vec{E}(x, y)$ which are related to the charge distribution through the Poisson equation

$$-\nabla^2 V(x, y) = \vec{\nabla} \cdot \vec{E}(x, y) = \frac{\rho(x, y)}{\epsilon} \quad (14)$$

It was shown by Büttiker¹⁵ and Entin-Wohlman et. al.¹⁶ that in the linear response regime, and in situations where the spatial variation in the potential $V(x, y)$ is smooth on the scale of the screening length (i.e. the screening is very strong), the potential is given by

$$V(x, y) = \frac{\sum_m \mu_1 |\psi_{1,m}(x, y)|^2 + \mu_2 |\psi_{2,m}(x, y)|^2}{\sum_m (|\psi_{1,m}(x, y)|^2 + |\psi_{2,m}(x, y)|^2)} \quad (15)$$

where μ_1 and μ_2 are the chemical potentials (Fermi energy) in the two contacts of the structure under investigation, $\psi_{1,m}$ is the wavefunction of electrons injected from the left contact in subband m with energy μ_1 and $\psi_{2,m}$ is the wavefunction of electrons injected in mode m with energy μ_2 from the right contact. It was pointed out by Entin-Wohlman et. al.¹⁶ that $eV(x, y)$ is also what is actually measured at a chemical potential probe so that it is also the chemical potential. In other words, in this regime, $eV \equiv \mu$.

Once the electrostatic (or chemical) potential is determined, the corresponding electric field is found simply from

$$\vec{E}(x, y) = -\vec{\nabla} V(x, y) = -\vec{\nabla} \mu(x, y) \quad (16)$$

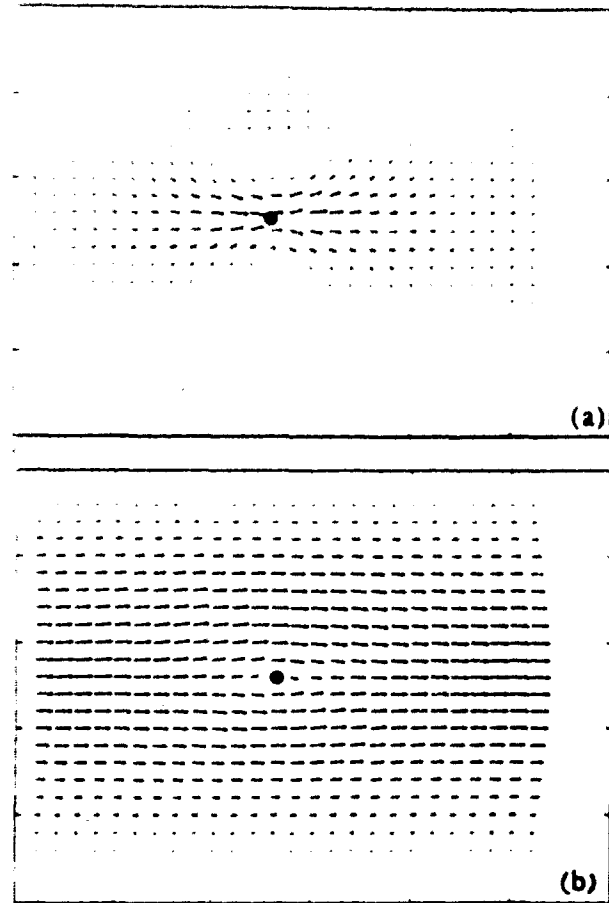
This is the electric field that should enter Equation (13) for the direct force in the linear response regime. Note that this means that the wind force and the direct force are equal in this regime if the valency Z of the ion is unity.

In the next section, we present results dealing with the spatial distribution of the current, carrier concentration change caused by the current, the chemical potential profile, the residual resistivity dipole potential, and the wind force and direct force on impurities in a prototypical GaAs quantum wire containing elastic delta scatterers. We consider both attractive and repulsive scatterers and show some typical examples of how the spatial patterns respond to an external magnetic field.

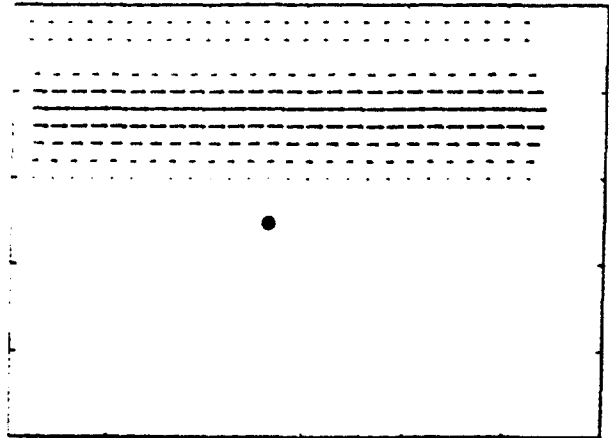
Results

Spatial distribution of the current around a single scatterer - local effects: We consider a GaAs wire which is 800 Å long and 1000 Å wide. The Fermi energy is 2.054 meV corresponding to an electron concentration of $3.85 \times 10^7/\text{m}$. For this low concentration, only one subband is occupied in both transverse directions so that transport is strictly single-channelled.

In Figures 3(a) and 3(b), we show the current profiles around a single attractive and repulsive scatterer respectively in the absence of any magnetic field. In the case of an attractive scatterer, the current is drawn closer to the scatterer, whereas for a repulsive scatterer, the current detours around the scatterer. This



• Fig. 3: The spatial distribution of the current density $J_{\text{total}}(x, y)$ in a 800 Å long and 1000 Å wide structure in the absence of any magnetic field. The Fermi energy is 2.054 meV. (a). Current crowding around an attractive impurity, (b), current detour around a repulsive impurity.



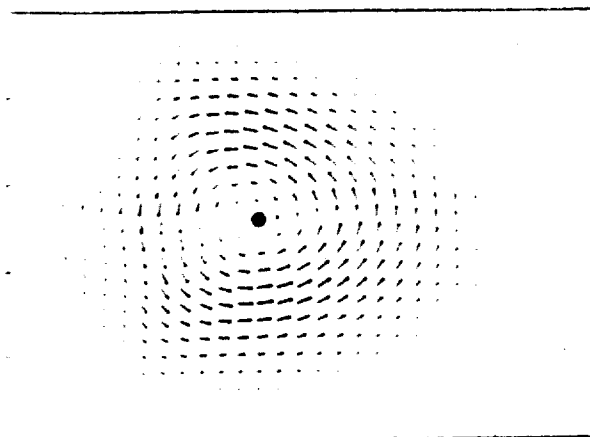
• Fig. 4: The spatial distribution of the current density $J_{\text{total}}(x, y)$ in the structure of Fig. 3(a) when a magnetic flux density of 2.18 Tesla is applied. The current flows along the top edge indicative of the formation of "edge states". This situation corresponds to the onset of the integer quantum Hall effect.

behavior is expected from electrostatic attraction and repulsion between electrons and the scatterer.

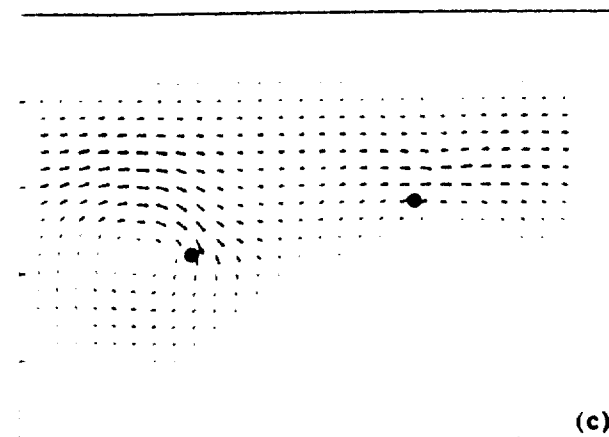
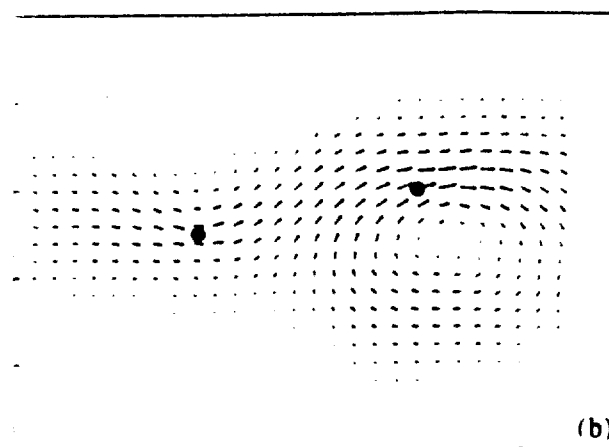
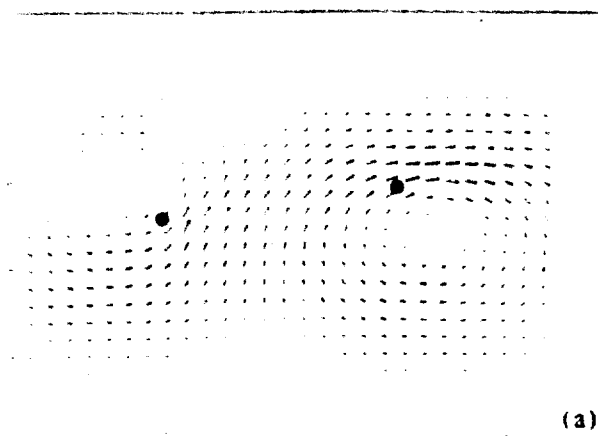
In Fig. 4, we show how the current pattern changes when a magnetic flux density of 2.18 Tesla is applied to the structure. The impurity is attractive. The current now flows entirely along one of the edges indicative of the formation of "edge states". There is practically no backscattering and the current flows straight through (the small reverse traveling component near the very top of the figure is not due to backscattering; it is due to skipping orbits and its origin was explained in reference 17). This current pattern is characteristic of the onset of the integer quantum Hall effect. The role of perfectly transmitting edge states in the quantization of Hall resistance was elucidated in reference 18.

In Fig. 5, we show the circulating current pattern around an attractive impurity when a magnetic bound state forms. The magnetic flux density is 3.5 Tesla. Note that the circulating current patterns hardly carry any net current in the x-direction. Therefore, the conductance of the structure will be very low when a magnetic bound state forms. It was shown in reference 1 that the conductance does indeed drop abruptly when a magnetic bound state forms.

Spatial distribution of the current around two scatterers - non-local quantum interference effects: In Fig. 6(a), we show the current pattern when two attractive impurities are present and no magnetic field is applied. Strong vortices form due to quantum interference between waves reflected from the two impurities. These vortices are purely a result of non-local quantum effects. The physics of these vortices were dealt with at length in the first two citations in reference 1. These vortices are interesting since they give rise to localized magnetic moments. In Fig. 6(b), we show how one of the vortices is destroyed at a flux density of 0.08 Tesla while the other remains. At a flux density of 1 Tesla, (Fig. 6(c)), the first vortex reappears.



• Fig. 5: The spatial distribution of the current in the structure of Fig. 3(a) when the magnetic flux density is increased to 3.5 Tesla. At this field a magnetic bound state forms around the impurity which causes the circulating current pattern. The net transmission of current through the structure is very low so that the conductance of the structure drops abruptly when a magnetic bound state forms.



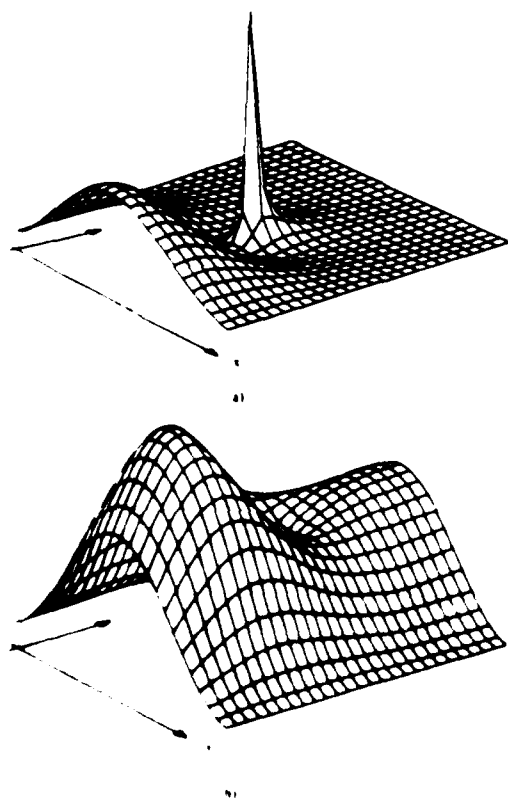
• Fig. 6: The spatial distribution of the current in a 900 \AA long and 1000 \AA wide structure containing two attractive impurities. (a). No magnetic field is present and two vortices form as a result of quantum interference between waves reflected from the two scatterers and the walls of the structure. (b). a magnetic flux density of 0.08 Tesla is applied and one of the vortices is quenched. (c). the magnetic flux density is increased to 1 Tesla and the first vortex reappears at a different location while the second vortex disappears.

appears at a different location and the second vortex is quenched with the simultaneous beginning of the formation of edge states.

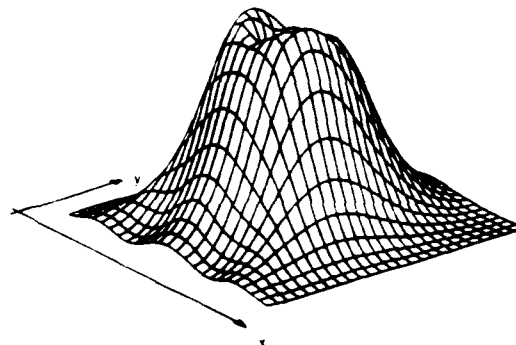
Carrier concentration change due to current flow: In Fig. 7(a) and 7(b), we show the change in the carrier concentration (from the equilibrium value) around the attractive and repulsive scatterers of Figs. 3(a) and 3(b) respectively. Carriers pile up around the attractive impurity and are depleted around the repulsive impurity because of electrostatic interaction.

In Fig. 8, we show the carrier pile up around the impurity when a magnetic bound state forms. The formation of such a state can cause significant inhomogeneities in the carrier profile and the associated electric field.

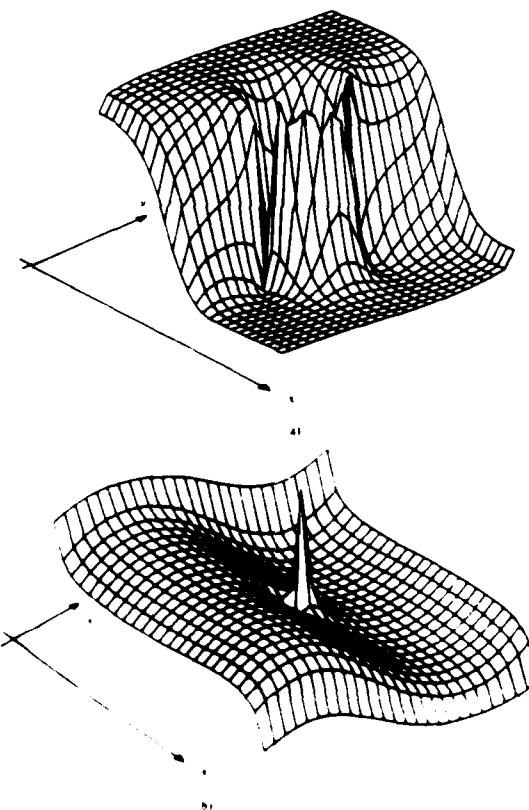
Chemical potential profile: In Figure 9(a), we show the chemical potential profile around the attractive impurity of Fig. 3(a). We assume that the chemical potentials at the two contacts of the structure (at $x=0$ and $x=L$) are $\mu_1 = 2.056$ meV and $\mu_2 = 2.052$ meV. Almost all of the chemical potential difference ($\mu_1 - \mu_2$) is dropped at the center where the impurity is located. This is expected since the impurity that is the major cause of the resistance for this sample. There are some oscillations in the chemical potential around the impurity but these are just signatures of quantum interference effects. The case of a repulsive impurity is very different. In this case (Fig. 9(b)), we see that



• Fig. 7 The spatial distribution of the change in the carrier concentration $\delta n_c(x, y)$ from the equilibrium value due to the current. No magnetic field is applied. (a) the scatterer is attractive. (b) the scatterer is repulsive.



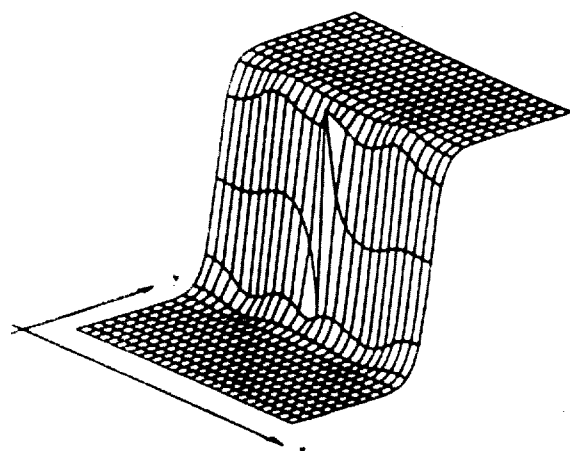
• Fig. 8: The spatial distribution of $\delta n_c(x, y)$ at a magnetic flux density of 3.5 Tesla when a magnetic bound state forms.



• Fig. 9 The spatial distribution of the chemical potential in the absence of any magnetic field. (a) the scatterer is attractive (corresponding to the case of Fig. 3(a)). (b) the scatterer is repulsive (corresponding to the case of Fig. 3(b)). Note that in the first case, most of the potential drop occurs around the impurity which is the dominant cause of resistance. In the second case, the potential is dropped more at the contacts showing that the contact resistances are dominant over the residual resistance of the impurity. This is a major difference between attractive and repulsive scatterers.

the potential drop occurs more near the contacts than around the impurity. There is only a narrow peak around the impurity but not much of a net drop from one end of the impurity to the other. This means that the major contributor for the resistance in this case is not the impurity, but the contact resistance. A repulsive impurity appears to be much less effective than an attractive impurity of the same scattering cross-section in reflecting electrons and causing resistance. We believe that this is related to the fact that quasi-donor states (quasi bound states) form around an attractive impurity¹⁹ but not around a repulsive one. When quasi donor states form, evanescent waves build up around the impurity which cause a tremendous amount of reflection and hence resistance. *This difference between the resistances caused by an attractive impurity and a repulsive impurity of the same scattering cross-section is a major difference between majority carrier transport (when most scatterers are attractive) and minority carrier transport (when most scatterers are repulsive).* Note that this difference is purely a consequence of quantum mechanics since semiclassical scattering theories, such as the Born approximation or Fermi's Golden rule, do not discriminate between attractive and repulsive scatterers. An experimental verification of this difference would be to demonstrate a large difference between majority carrier mobilities (measured by Hall effect) and minority carrier mobilities (measured by Shockley-Haynes method).

In Figure 10, we show how the chemical potential profile of Figure 9(a) changes when a magnetic flux density of 2.18 Tesla is turned on. The edges of the wire (along the length) become strikingly smooth equipotential surfaces. Consequently, the longitudinal resistance measured by attaching two voltage probes at any of the two edges will read exactly zero since the potential

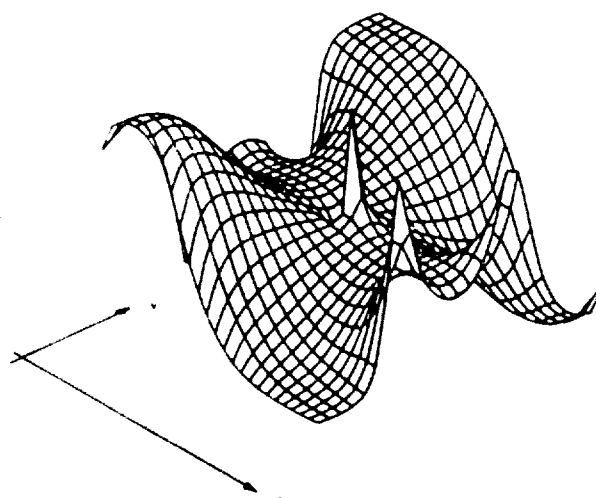


• Fig. 10: The spatial distribution of the chemical potential at a magnetic flux density of 2.18 Tesla. Note that the edges of the structure (along the x-direction) are exceptionally smooth equipotential surfaces so that the longitudinal resistance measured by attaching two probes at either edge will be exactly zero since the chemical potential difference between these two probes will be zero. The potential drop along the y-direction (Hall voltage) is independent of the x-coordinate and is exactly equal to the potential drop between the two contacts of the structure. The Hall resistance is exactly $h/2e^2$. This figure is a direct visualization of the integer quantum Hall effect.

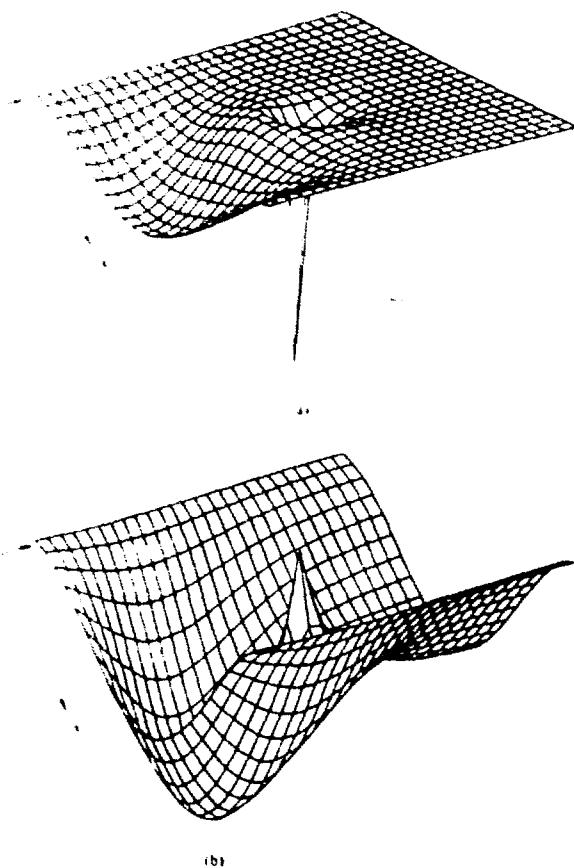
difference measured by these probes will be exactly zero. This is a manifestation of the *quantum Hall effect*. Also the Hall voltage drop (the difference between the chemical potentials at the two edges) is found to be exactly $\mu_1 - \mu_2$. Therefore, the Hall resistance is $(\mu_1 - \mu_2)/I$ where I is the current flowing in the structure. The current I is given by $I = G(\mu_1 - \mu_2)$, where G is the two-terminal conductance. In reference 1, we calculated G for this structure and found it to be precisely $2e^2/h$. Therefore, the Hall resistance is exactly $h/2e^2$ which is again a manifestation of the integer quantum Hall effect.

In Fig. 11, we show how the chemical potential changes when the flux density is increased to 3.5 Tesla. At this flux density, a magnetic bound state forms. The bound state couples electrons between the two sets of edge states carrying current along the two edges of the wire and causes backscattering. This destroys the quantum Hall effect²⁰ - a fact that is clearly seen from the chemical potential profile. The edges of the sample are no longer equipotential surfaces so that the longitudinal resistance will no longer measure zero. Also, the Hall voltage drop is different from $(\mu_1 - \mu_2)$ so that the Hall resistance is not quantized to sub-multiples of h/e^2 . *This is an unequivocal demonstration that a magnetic bound state in a narrow wire destroys the integer quantum Hall effect.*

Spatial distribution of the residual resistivity dipole potential: In Figs. 12(a) and 12(b), we show the residual resistivity dipole potential in the case of the single attractive scatterer and the single repulsive scatterer of Figs. 3(a) and 3(b). The potential has an approximately $1/r$ decay or rise (where r is the radial distance from the scatterer), as predicted by Chu and Sorbello for the case of a two-dimensional electron gas².



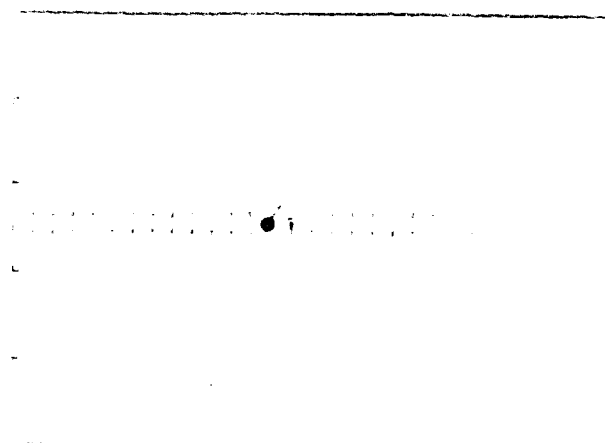
• Fig. 11: The spatial distribution of the chemical potential when the magnetic flux density is increased to 3.5 Tesla and a magnetic bound state forms. The edges are no longer equipotential surfaces and the Hall voltage drop is very different. The longitudinal resistance is no longer zero and the transverse resistance is no longer quantized to sub-integral multiples of h/e^2 . This figure is a direct visualization of the destruction of the integer quantum Hall effect in a narrow wire by a magnetic bound state.



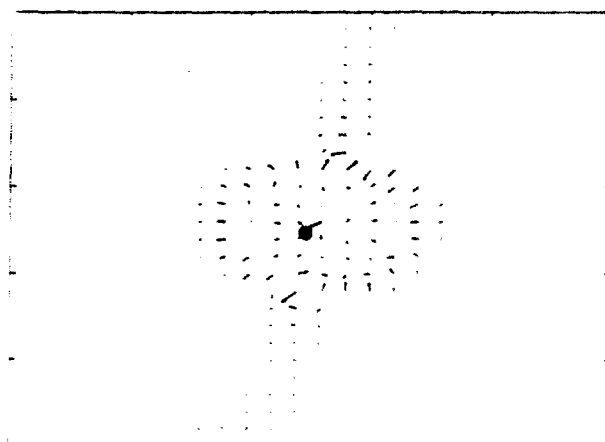
• Fig. 12: The spatial distribution of the residual resistivity dipole potential in the absence of any magnetic field. (a), the scatterer is attractive, (b), the scatterer is repulsive.

Spatial distribution of the electric field: In Fig. 13, we show the electric field distribution in the quantum Hall regime corresponding to the situation in Fig. 10. All the field is concentrated at the center of the wire and is basically the Hall field. When the magnetic field is turned off, the electric field profile changes to that shown in Fig. 14. There is no net field in the y -direction since there is no Hall voltage drop in this case.

The forces of electromigration: In Figs. 15(a) and 15(b) we show the directions of the wind forces and also the direct forces (recall that the two forces are equal) on two impurities in a 900 \AA long and 1000 \AA wide structure at zero magnetic field and at a magnetic flux density of 3.5 Tesla. Unlike in the case of a single impurity, no magnetic bound state forms at the flux density of 3.5 Tesla when two impurities are present. Note that at zero magnetic field, the wind forces on neither impurity is in the direction of the electric field (x -direction) in contrast to the predictions of semiclassical theories. This difference in the orientation of the forces is a quantum mechanical effect. In addition, magnetic field can drastically change the magnitude and the direction of the electromigration forces on both impurities. In the case of the first impurity on the left, the magnetic field almost completely removes the forces. In the case of the second impurity, the forces are enhanced by the magnetic field. Therefore, in the quantum coherent regime, whether a magnetic field



• Fig. 13: The spatial distribution of the electric field in the quantum Hall regime corresponding to the situation depicted in Fig. 10. The magnetic flux density is 2.18 Tesla. All the field is at the center and is directed along the y -direction (Hall field).

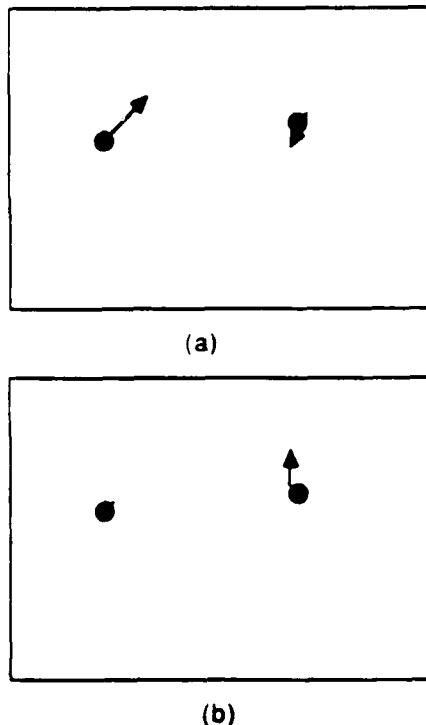


• Fig. 14: The spatial distribution of the electric field when the magnetic field is turned off.

increases or decreases the direct force on an ion (or whether it has any effect at all), is determined by the *locations* of the ions within the sample. This is purely a quantum-mechanical phenomenon.

Conclusions

In this paper, we have presented the spatial distributions of a large number of transport variables associated with various quantum magnetotransport phenomena in mesoscopic samples. These distributions elucidate the integer quantum Hall effect, the formation of magnetic bound states and its effect on the integer quantum Hall effect, the difference between resistances caused by attractive and repulsive scatterers of the same scattering cross-section and the corresponding difference between minority and majority carrier mobilities, the magnetic response of current vortices formed as a result of quantum interference between scatterers, the electric field distribution in the quan-



• Fig. 15: The forces of electromigration acting on two impurities in a 900 Å long and 1000 Å wide structure. (a). No magnetic field is present. (b). The magnetic flux density is 3.5 Tesla.

turn Hall regime, the residual resistivity dipole potential and its spatial variations, and finally, the direct and wind forces of electromigration and their dependences on a magnetic field as well as the location of the ion. We believe that these results will be useful in interpreting many different types of transport phenomena and electromigration in mesoscopic structures.

Acknowledgement - This work was supported by the US Air Force Office of Scientific Research under Grant No. AFOSR 91-0211, by the US Office of Naval Research under contract no. N00014-91-J-1505, and by the National Science Foundation under grant no. ECS 9108932.

REFERENCES

1. S. Chaudhuri, S. Bandyopadhyay and M. Cahay, *Physical Review B*, Vol. 45, 11126 (1992); S. Chaudhuri, S. Bandyopadhyay and M. Cahay, to appear in *Physical Review B*; C. S. Lent, *Applied Physics Letters*, Vol. 67, 1678 (1990); Zhen-Li Ji and K. F. Berggren, *Physical Review B*, Vol. 45, 6652 (1992); K. F. Berggren, C. Bozov and Zhen-Li Ji, to appear in *Physica Scripta*; Zhen-Li Ji, *Semiconductor Science and Technology*, Vol. 7, 198 (1992).
2. R. S. Sorbello, *Physical Review B*, Vol. 23, 5119 (1981); R. S. Sorbello and C. S. Chu, *IBM Journal of Research and Development*, Vol. 32, 58 (1988).
3. R. Landauer, *IBM Journal of Research and Development*, Vol. 1, 223 (1957); R. Landauer and J. W. F. Woo, *Physical Review B*, Vol. 10, 1266 (1974); R. Landauer, *Z. Physik B*, Vol. 21, 247 (1975).
4. S. Chaudhuri and S. Bandyopadhyay, *Journal of Applied Physics*, Vol. 71, 3027 (1992).
5. V. B. Fiks, *Soviet Physics - Solid State*, Vol. 1, 14 (1959); H. B. Huntington and A. R. Grone, *Journal of Physics and Chemistry of Solids*, Vol. 20, 76 (1960).
6. There are subtle issues regarding whether the electron momentum should be taken as the crystal momentum or the real momentum or a combination of the two. This question has not been resolved satisfactorily and we will assume for the sake of simplicity that the relevant momentum is the crystal momentum. Since, in this paper, we will only calculate the direction and not the exact magnitude of the wind force, we believe that the distinction between crystal momentum and real momentum is not very important in this context.
7. S. Bandyopadhyay, M. E. Klausmeier-Brown, C. M. Maziar, S. Datta and M. S. Lundstrom, *IEEE Transactions on Electron Devices*, Vol. ED-34, 392 (1987).
8. C. Bosvieux and J. Friedel, *Journal of Physics and Chemistry of Solids*, Vol. 23, 123 (1962).
9. R. S. Sorbello, *Journal of Physics and Chemistry of Solids*, Vol. 34, 937 (1973); A. H. Verbruggen, *IBM Journal of Research and Development*, Vol. 32, 93 (1988).
10. R. P. Gupta, Y. Serruys, G. Brebec and Y. Adda, *Physical Review B*, Vol. 27, 672 (1983).
11. A. K. Das and R. Peierls, *Journal of Physics C: Solid State Physics*, Vol. 6, 2811 (1973).
12. R. Landauer, *Journal of Physics C: Solid State Physics*, Vol. 8, 761 (1975); R. Landauer, *Physical Review B*, Vol. 14, 1474 (1976).
13. H. B. Huntington, *Transactions of Metallurgical Society*, Vol. AIME 245, 2571 (1969).
14. L. Lou, W. L. Schaich and J. C. Swihart, *Physical Review B*, Vol. 33, 2170 (1986).
15. M. Büttiker, *IBM Journal of Research and Development*, Vol. 32, 317 (1988).
16. O. Entin-Wohlman, C. Hartsztein and Y. Imry, *Physical Review B*, Vol. 34, 921 (1986).
17. C. S. Lent, *Physical Review B*, Vol. 43, 4179 (1991).
18. M. Büttiker, *Physical Review B*, Vol. 38, 9375 (1988).
19. P. F. Bagwell, *Physical Review B*, Vol. 41, 10354 (1990).
20. J. K. Jain and S. Kivelson, *Physical Review Letters*, Vol. 60, 1542 (1988); Y. Lee, M. J. McLennan and S. Datta, *Physical Review B*, Vol. 43, 14333 (1991).

Current, potential, electric field, and Fermi carrier distributions around localized elastic scatterers in phase-coherent quantum magnetotransport

S. Chaudhuri and S. Bandyopadhyay

Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

M. Cahay

Department of Electrical and Computer Engineering, University of Cincinnati, Cincinnati, Ohio 45221

(Received 3 April 1992; revised manuscript received 28 December 1992)

The spatial distributions of the current, Fermi carriers, electric field (due to space charges) and electrostatic potential in a disordered mesoscopic structure are calculated in the presence of arbitrary magnetic fields. These distributions are useful in elucidating and visualizing many features of quantum magnetotransport, such as the formation of edge states at high magnetic fields and their near-perfect transmittivity, the evolution of the integer quantum Hall effect, the creation of magnetic bound states around an impurity, the magnetic response of current vortices that form as a result of quantum interference between scatterers and the walls of a quantum wire, the dependence of the quantized-conductance steps in a backgated quantum wire on an applied magnetic field, the behavior of residual-resistivity dipoles and the electrostatic space-charge potential in a magnetic field, the dependence of the sign of the magnetoresistance on the impurity configuration, etc. We examine the current, Fermi carrier concentrations, electric field, and both chemical- and electrostatic-potential profiles associated with each of these phenomena, and relate them to the observed terminal characteristics in each case.

I. INTRODUCTION

Many features of quantum magnetotransport in a mesoscopic structure can be understood by examining the spatial distributions of the current, potential, electric field, and Fermi carriers around elastic scatterers in the presence of a magnetic field. In this paper, we have calculated such spatial patterns from a fully microscopic quantum-mechanical formalism based on the Schrödinger equation. These spatial distributions are, of course, not "precise" in view of the Heisenberg uncertainty principle which prohibits simultaneous definition of both the position and the momentum of an electron; instead they should be viewed as spatial distribution of quantum-mechanical averages (or expected values) of the corresponding quantities. These patterns help in understanding several quantum transport phenomena and provide direct visualization of the associated physics. For example, we show the spatial distribution of the current carried by edge states in a quantum wire at high magnetic field. The current patterns clearly show the high transmittivity which is responsible for the quantization of the Hall resistance in the quantum Hall regime. We also show how current circulates around an impurity when a magnetic bound state forms. The recovery of the quantization of conductance steps in a dirty quantum wire by the application of a magnetic field is known to be associated with the suppression of backscattering, and the current patterns that we obtain show this very clearly. The onset of the integer quantum Hall effect is caused by the edges of a quantum wire becoming perfectly smooth equipotential surfaces and our chemical potential profiles demonstrate this strikingly. The magnetic response of current vortices, electrostatic potential, and electric field

all show quenching phenomena. We also found why the sign of the magnetoresistance of a quantum wire could depend on the position of a single impurity inside the wire. These, and the nature of other phenomena, become very clear when one examines the spatial patterns of current, carrier concentration, and electrostatic or chemical potentials. The spatial distributions are not only of great importance in understanding the origin and the detailed physics of many quantum magnetotransport phenomena, but they often determine the nature of certain physical phenomena such as electromigration.

This paper is organized as follows. In Sec. II, we briefly present the quantum-mechanical theory used in computing the two-dimensional spatial distributions of current, Fermi carrier concentration, potential, and internal electric-field profiles around localized scatterers in the presence of a magnetic field. In Sec. III, we present the results for a disordered GaAs quantum wire containing a variable number of impurities. We show the spatial patterns associated with various magnetotransport phenomena in both single-channeled and multichanneled transport. Such spatial patterns were presented by us in the limit of coherent diffusive transport (in the absence of any magnetic field) earlier.¹ Others have presented such patterns in the limit of coherent ballistic transport with no scattering whatsoever.² To our knowledge, we are the first to present such patterns in the *coherent diffusive regime in the presence of a magnetic field*.

II. THEORY

The theory for our calculations of the current and carrier density patterns in a quantum wire in the absence of a magnetic field has been described in Ref. 1. Here, we

repeat only the basic features. We consider a mesa as shown in Fig. 1 which is single moded in the z direction (i.e., only a single subband is occupied in that direction), but multimoded in the y direction.

The time-independent Schrödinger equation describing steady-state electron transport in this confined quasi-one-dimensional disordered structure under a magnetic field is

$$\frac{(\mathbf{p} - e \mathbf{A})^2}{2m^*} \psi + V\psi = E\psi, \quad (1)$$

where \mathbf{p} is the momentum operator, m^* is an isotropic effective mass, and the potential energy V consists of two terms,

$$V \equiv V(x, y) = V_c(y) + V_{\text{imp}}(x, y). \quad (2)$$

The first term gives the effect of the confining potential in the y direction and the second term is due to impurities. We assume hardwall boundary conditions in the y direction and δ potentials for the impurities so that we have

$$V_{\text{imp}}(x, y) = \gamma \sum_{i=1}^s \delta(x - x_i) \delta(y - y_i), \quad (3)$$

where s is the total number of impurities in the structure and the summation extends over all the impurities. The position of the i th impurity is denoted by (x_i, y_i) .

The wave function ψ^0 in the absence of impurities can be written as

$$\psi^0(x, y) = e^{ik_x x} \phi(y), \quad (4)$$

where k_x is the x -directed wave vector and $\phi(y)$ satisfies the eigenequation

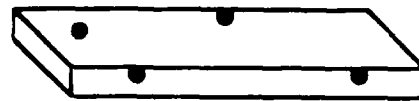


FIG. 1. A quasi-one-dimensional quantum wire containing a random array of elastic scatterers. The structure is single moded in the z direction but multimoded in the y direction. The magnetic field is applied in the z direction.

$$\frac{\partial^2 \phi}{\partial y^2} + \frac{2m^*}{\hbar^2} E \phi(y) - \left[\frac{y}{l^2} \right]^2 \phi(y) + 2 \frac{y}{l^2} k_x \phi(y) - k_x^2 \phi(y) = 0, \quad (5)$$

with l being the magnetic length given by $l = \sqrt{\hbar/eB}$.

The wave function $\phi(y)$ is found by solving the above equation numerically using a finite difference scheme as described in Ref. 3. This method also gives the energy dispersion relations E vs k_x for the various hybrid magnetoelectric subbands that arise from the confining effects of the magnetic field and the confining potential in the y direction. A set of computed energy dispersion relations are shown in Fig. 2. After calculating the energy dispersion relations, we choose a certain energy E (which we call the Fermi energy E_F) and find the corresponding x -directed wave vectors k_p^F for various magnetoelectric subbands (indexed by p) from the dispersion relations. These wave vectors k_p^F and the wave functions $\phi_p(y)$ are used to determine the total wave function $\psi_m(x, y)$ (in the presence of impurities) for an electron injected with the Fermi energy E_F from the subband m in the left contact. The wave function is given by

$$\psi_m(x, y) = \sum_{p=1}^M \frac{A_{mp}(x) \exp(ik_p^F x) \phi_p(y) + B_{mp}(x) \exp(ik_{-p}^F x) \phi_{-p}(y)}{\sqrt{|v_p^F|}}, \quad (6)$$

where the summation extends over both propagating and evanescent modes (the total number of modes considered is M). The subscript $-p$ indicates that the quantity corresponds to a wave function with an oppositely directed velocity as compared to the one with a subscript p .

When the wave function is written in this form (i.e., each mode is normalized by the square root of the corresponding group velocity), all propagating modes at the Fermi energy carry the same current. This is consistent with the well-known cancellation of the velocity and density of states in quasi-one-dimensional structures which causes every mode to carry the same current.⁴ We now

only have to evaluate the coefficients $A_{mp}(x)$ and $B_{mp}(x)$ to determine the wave function of Fermi electrons everywhere. If we know these coefficients at any position (x_0, y) in the structure, then the coefficients at any other position (x_1, y) can be found from

$$\begin{bmatrix} r_0 \\ r_1 \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{bmatrix} \begin{bmatrix} l_1 \\ l_0 \end{bmatrix}, \quad (7)$$

where l_i is a column vector of length M whose elements are $A_{m1}(x_0)$, $A_{m2}(x_0)$, $A_{m3}(x_0)$, ..., $A_{mM}(x_0)$, l_0 is a

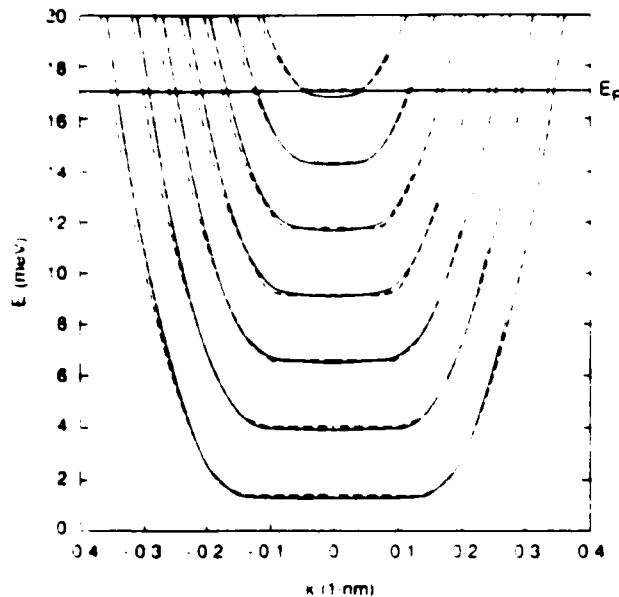


FIG. 2. The energy dispersion relations for the hybrid magnetoelectric subbands in a quantum wire of width 2000 Å at a magnetic flux density of 1.5 T. The solid lines are the results of exact numerical calculations, whereas the broken lines are the results of the semiclassical Bohr-Sommerfeld quantization rule.

column vector of length M whose elements are $B_{m1}(x_0), B_{m2}(x_0), B_{m3}(x_0) \dots B_{mM}(x_0)$, r_0 is a column vector of length M whose elements are $A_{m1}(x_1), A_{m2}(x_1), A_{m3}(x_1) \dots A_{mM}(x_1)$, and r_1 is also a column vector of length M whose elements are $B_{m1}(x_1), B_{m2}(x_1), B_{m3}(x_1) \dots B_{mM}(x_1)$. The coefficients t_{11}, t_{12}, t_{21} , and t_{22} are each $M \times M$ matrices and the square matrix in Eq. (7) is the so-called transfer matrix for the section of the structure between $x = x_0$ and x_1 .

The above equation allows us to determine the coefficients $A_{mp}(x_1)$ and $B_{mp}(x_1)$ (and hence the wave function) at any arbitrary location x_1 if we know the wave function [or the coefficients $A_{mp}(x_0)$ and $B_{mp}(x_0)$] at any one point $x = x_0$ in the device as well as the transfer-matrix elements. We do know the wave function at the left contact ($x = 0$). We assume that for injection into the m th mode, the wave function is given by scattering states⁵ (this is our boundary condition)

$$\psi_m(0-, y) = \frac{\exp(ik_m^F x) \phi_m(y)}{\sqrt{|v_m^F|}} + \sum_{p=1}^M \frac{R_{mp} \exp(ik_{-p}^F x) \phi_{-p}(y)}{\sqrt{|v_p^F|}} \quad (8)$$

so that $A_{mp}(x=0-) = \delta_{mp}$ (Kronecker delta) and $B_{mp}(x=0-) = R_{mp}$. Therefore, we can find the wave function anywhere in the structure using Eqs. (6) and (7) if we can find the elements of the appropriate scattering matrices in Eq. (7). In the absence of a magnetic field, these elements are fairly easy to find,¹ but for nonzero magnetic fields, the wave functions of the various subband states at the Fermi energy are *not* orthogonal. Con-

sequently, the elements of the scattering matrix cannot be found by the normal prescription of Ref. 1. The method for finding these elements is described in the Appendix.

Once we have evaluated the scattering matrices, we have found all the quantities that we need in order to evaluate the wave function of Fermi electrons everywhere in the structure. From this wave function, we calculate all quantities of interest.

In linear-response transport at low temperatures, only electrons at the Fermi level carry a net current. Therefore, the two-dimensional wave function $\psi_m(x, y)$ of electrons in the m th subband allows us to calculate the contribution of that subband to the two-dimensional current density. This is given by

$$J_m(x, y) = \frac{iq\hbar}{2m^*} [\psi_m(x, y) \nabla \psi_m^*(x, y) - \psi_m^*(x, y) \nabla \psi_m(x, y)] + \frac{1}{m^*} (e A |\psi_m|^2). \quad (9)$$

The Fermi carrier concentration due to the contribution from the m th subband is defined by us as

$$n_{im}(x, y) = |\psi_m(x, y)|^2. \quad (10)$$

The total current density is found by vectorially adding the contributions from all propagating modes and the total Fermi carrier concentration is found by the scalar addition of the contributions from all propagating modes. Note that no weighting by the density of states is necessary in the summation since the wave functions have been normalized by the square root of velocity which takes care of the density of states.

$$J_{\text{total}}(x, y) = \sum_{\text{propagating modes}} J_m(x, y), \quad (11)$$

$$n_i^{\text{total}}(x, y) = \sum_{\text{propagating}} n_{im}(x, y).$$

Finally, we find the electrostatic space-charge potential $V(x, y)$ from the relation^{4,6,7}

$$eV(x, y) = \frac{\sum_{\text{propagating}} |\psi_{1,m}(x, y)|^2 \mu_1 + |\psi_{2,m}(x, y)|^2 \mu_2}{\sum_{\text{propagating}} |\psi_{1,m}|^2 + |\psi_{2,m}|^2}, \quad (12)$$

where μ_1 and μ_2 are the chemical potentials in the two contacts of the quantum wire, $\psi_{1,m}$ is the wave function corresponding to injection in mode m (with energy μ_1) from the left contact, and $\psi_{2,m}$ is the wave function for injection in mode m (with energy μ_2) from the right contact. The above equation is derived from the Poisson equation and is valid only for (a) linear-response transport, (b) situations where the potential variation of $V(x, y)$ is smooth on the scale of the screening length, and (c) when the Fermi wavelength is considerably smaller than the screening length. When these conditions are not satisfied, the actual Poisson equation must be solved simultaneously with the Schrödinger equation, rather than using the simplified equation (12). That is numerically much more demanding because in the Poisson equa-

tion we have to account for carriers at all energies, not just the Fermi energy. This means that we must solve the Schrödinger equation for all energies which is a more demanding task than the present approach of solving it only at the Fermi energy. Fortunately, in semiconductors and in linear-response transport, the present approach may work well since space-charge effects are not tremendously important.⁸

It was pointed out by Entin-Wohlman, Hartzstein, and Imry⁷ that $eV(x,y)$ is also what is actually measured at a chemical potential probe. In reality, the electrostatic potential is the convolution of the chemical potential with a screening function.⁸ In the limit of strong screening (δ impurities), this screening function approaches a δ function so that the electrostatic and chemical potential become equivalent.

Once the electrostatic potential is determined, the corresponding electric field is found simply from

$$\mathcal{E}(x,y) = -\nabla V(x,y). \quad (13)$$

In the next section, we present the spatial distributions of the current, Fermi carrier concentration, electrostatic (or chemical) potential, and electric-field patterns associated with various magnetotransport phenomena.

III. RESULTS

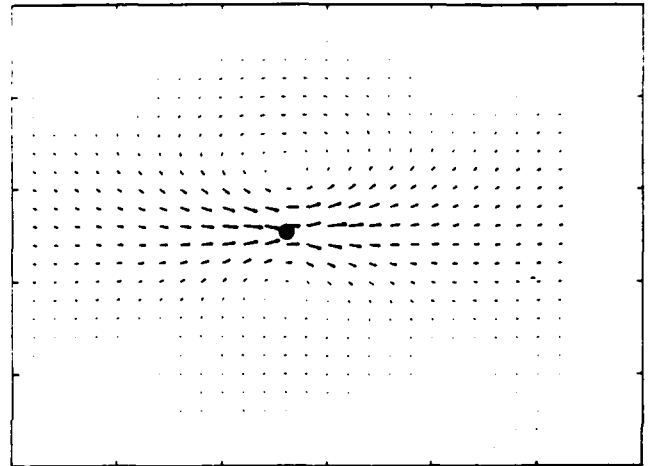
A. Spatial distributions around a single impurity in single-channeled magnetotransport: Local effects

In Figs. 3(a) and 3(b), we show the spatial distribution of the current around a single scatterer placed in the middle of a GaAs quantum wire in the absence of any magnetic field. The two figures are for an attractive and a repulsive scatterer, respectively. The wire is 800 Å long and 1000 Å wide. The Fermi energy is 2.054 meV and only a single subband is occupied in both y and z directions so that transport is single channeled. There is pronounced current crowding near the impurity when the impurity potential is attractive and diverges away from the impurity when the impurity potential is repulsive. There are also two weak vortices above and below the scatterer for an attractive impurity. If the impurity is positioned close to the edges of the wire instead of at the center, such a pronounced effect does not occur. There are two reasons for this. First, the wave function diminishes in amplitude near the edges of the wire so that the interaction between the electron and the impurity is weakened when the impurity is closer to the walls. Second, the confining potential of the walls plays a greater role nearer the walls so that the current sees not only the impurity, but also the walls. As a result, the current pattern is determined by the combined interaction of the impurity and the wall.

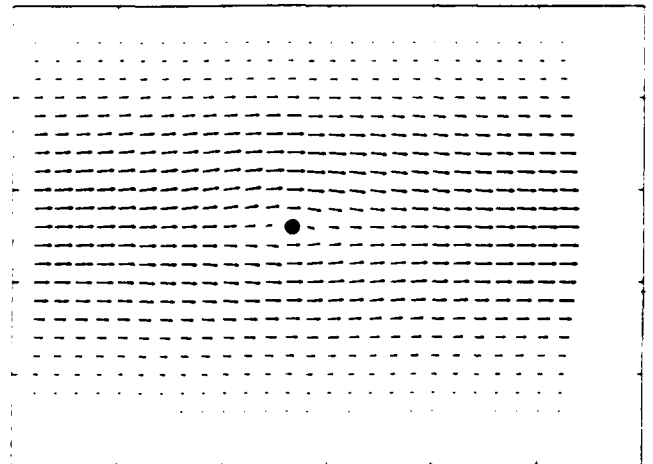
It is obvious that when the impurity is right at the center, the interaction between the impurity and the electrons is strongest since the wave function (of the lowest subband) peaks at the center. A magnetic field will skew the wave function away from the impurity and hence decrease the interaction or scattering. This will lead to a decrease in the resistance so that we will observe a nega-

tive magnetoresistance in this case. On the other hand, if the impurity is close to one of the walls, then a magnetic field will either bring the electrons closer to the impurity or take it further away depending on the direction of the magnetic field. Therefore we can observe either positive or negative magnetoresistance in that case (depending on the field direction) and the magnetoresistance will be asymmetric in the magnetic field. This is a rather striking effect which might be observable in mesoscopic structures at very low temperatures when impurity scattering is the dominant mechanism for resistance.

In Figs. 4(a) and 4(b), we have plotted the Fermi carrier concentration around the attractive and the repulsive



(a)



(b)

FIG. 3. The current distribution inside an 800-Å-long and 1000-Å-wide structure in the absence of any magnetic field. (a) The impurity is attractive and there is significant current crowding around the impurity which is shown by the solid circle. (b) The impurity is repulsive and the current detours around the impurity. This figure shows local effects of the impurity potential.

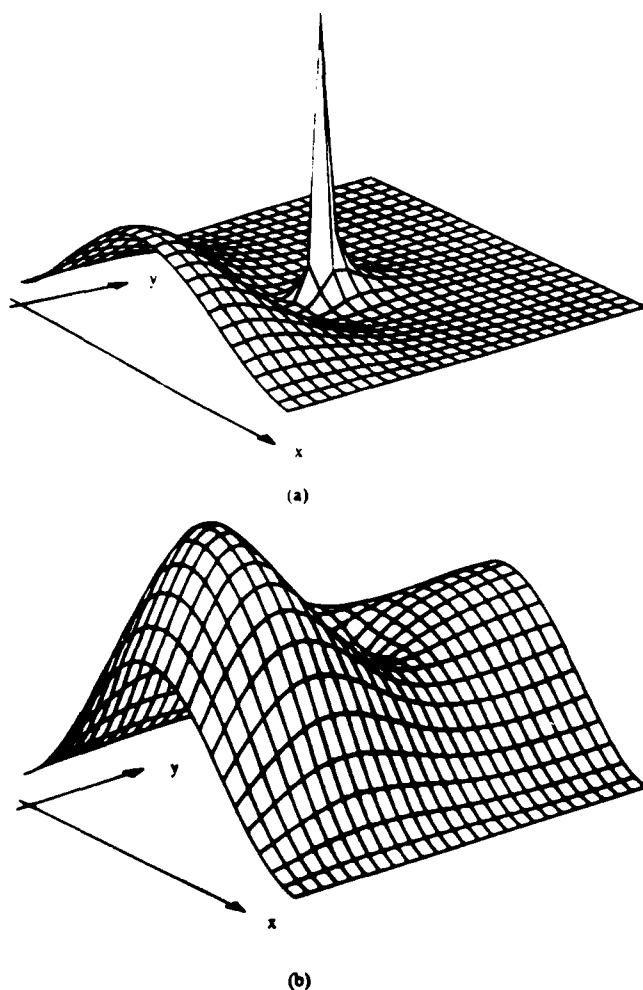


FIG. 4. The Fermi carrier concentration profile for the structure of Fig. 3 in the absence of any magnetic field. (a) The impurity is attractive; (b) the impurity is repulsive.

impurity of Figs. 3(a) and 3(b) in the absence of any magnetic field. As expected, in the case of the attractive impurity, we see a sharp spike (accumulation of electrons) around the impurity where the current crowding occurs. The electrons are attracted to the impurity by electrostatic attraction. In the case of repulsive impurity, the concentration is depleted where the impurity is positioned because of electrostatic repulsion. In Fig. 4(a), we see that there is a building up of electrons to the left of the scatterer and a deficit to the right. This happens because the impurity strongly reflects the incoming electrons which build up to the left. Because of the excess negative charge on one side and a deficit on the other, we expect a dipole to form. This is the residual-resistivity dipole as discussed by Landauer.⁹ We can contrast these figures with the profiles that were presented in Ref. 1 which considered only multiple-impurity systems. When a large number of impurities are placed inside the wire, the local charge accumulation and depletion effect that causes a dipole to form is usually diminished because of quantum-mechanical interference between waves reflected from the various impurities. This quantum-mechanical interference is a nonlocal effect which tends to inhibit local di-

pole formation in most cases, although it may sometimes also enhance the dipole formation depending on the impurity configuration, i.e., the exact phase relationships between the interfering waves. Therefore, even though the dipole formation is a local effect in itself, the size and orientation of the dipoles are significantly affected by nonlocal quantum interference effects.

In Figs. 5 and 6, we plot the electrostatic (or chemical) potential and electric-field profiles. For these figures, we assume that $\mu_1 = 2.056$ meV and $\mu_2 = 2.052$ meV. We do not exactly see the characteristic electric-field distribution of a dipole since the "poles" of the dipole are not "isolated" charges; they are "extended" charges. The plots are presented for both attractive and repulsive scatterers. They are very different when the impurity potential is attractive (which corresponds to majority carrier transport) as opposed to the case when it is repulsive (minority carrier transport). In the case of attractive impurities, there can be a building up of evanescent states around the impurities associated with the formation of quasibound donor states.¹⁰ These may help in dipole formation.

The application of a magnetic field causes dramatic changes in the current, carrier concentration, electrostatic potential, and electric-field profiles. We show these

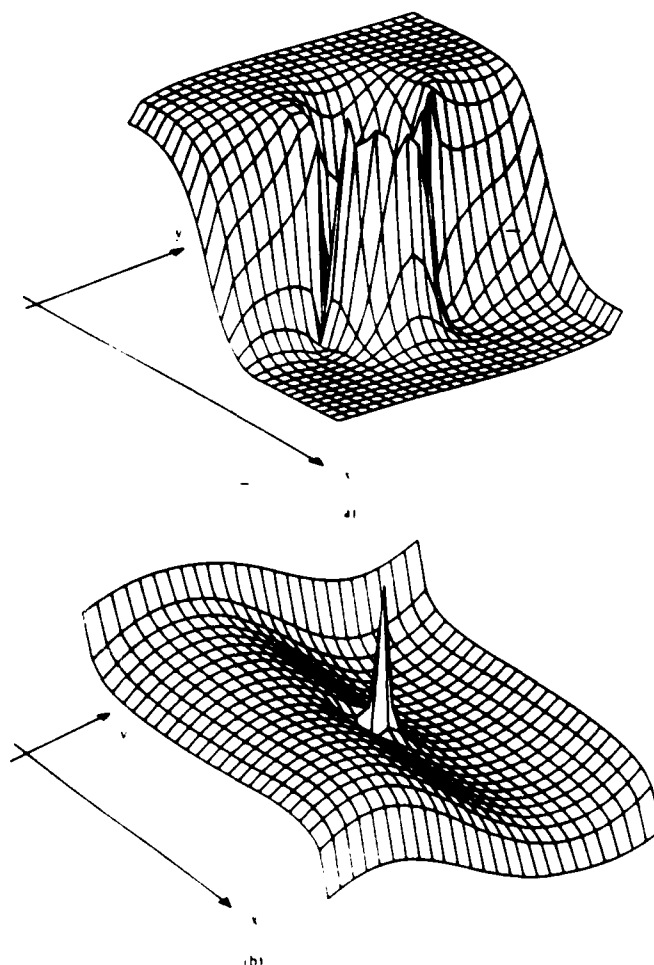


FIG. 5. The electrostatic- (or chemical-) potential profile inside the structure of Fig. 3 in the absence of any magnetic field. (a) Attractive impurity; (b) repulsive impurity.

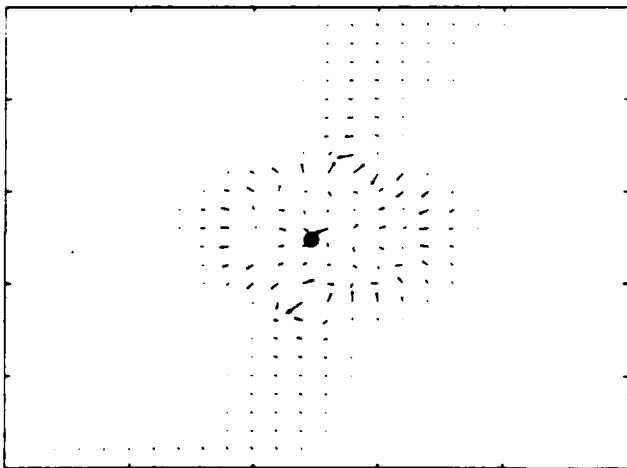
effects for the single attractive impurity in Figs. 7(a), 7(b), 7(c), and 7(d) which are to be compared with Figs. 3(a), 4(a), 5, and 6, respectively.

In Fig. 7(a), we have plotted the current inside the wire in the presence of a magnetic-flux density of $B = 2.18$ T. All other conditions correspond to the case of Fig. 3(a). Comparing with Fig. 3(a), we find that current crowding has disappeared and the two weak vortices have been quenched. The sample now no longer exhibits localized magnetic moments due to the vortices. In addition, there is a significant reduction of the backscattered current component. A similar reduction in backscattering is observed for a repulsive impurity also, except that it takes a lower magnetic field to reduce the backscattering by the same amount. The current pattern in Fig. 7(a) clearly shows that edge states have formed and carry the bulk of

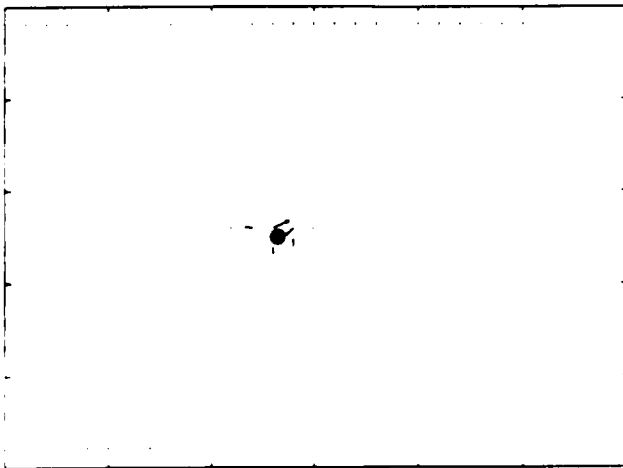
the forward current. There is also, however, a small reverse-traveling current component close to the walls. This component was explained in Ref. 11 (which dealt with edge states in ballistic transport as opposed to diffusive transport) as being due to the classical cyclotron trajectories (skipping orbits) having their velocities directed opposite to the net current flow adjacent to the edges. At higher magnetic fields, this reverse component diminishes so that backscattering is suppressed even more. In other words, the transmittivity of the edge states are very high at high magnetic fields. This agrees with Büttiker's picture of the integer quantum Hall effect¹² in which perfect transmission of edge states was invoked to explain quantization of the Hall resistance and vanishing of the longitudinal resistance.

In Fig. 7(b), we show the concentration profile of the Fermi carriers in the presence of a magnetic field of 2.18 T. The concentration piles up near an edge because of the Lorentz force skewing the wave function towards that edge. Note that there is no longer a buildup of electrons to the left of the scatterer and a deficit to the right since the buildup is due to reflection (backscattering) and this has diminished significantly. In fact, at high enough magnetic fields, backscattering is always reduced. Therefore, in general, residual-resistivity dipoles will be destroyed by an external magnetic field if it is of high enough magnitude.

In Fig. 8(a), we show the conductance of the structure of Fig. 3(a) versus the Fermi energy at a magnetic field of 3.5 T. There is only one attractive impurity in the structure placed at the center. A pronounced dip in the conductance is seen at an energy of 3.82 meV. At this energy, a magnetic bound state¹³ forms around the impurity which causes the transmission (and hence the conductance) to decrease sharply. This state forms at about the same energy for a repulsive scatterer. The corresponding current profile is shown in Fig. 8(b). We find that there is only a small amount of net current traveling from one contact to another when a magnetic bound state forms and the bulk of the current circulates around the impurity. These magnetic bound states are quite different from the quasidonor states¹⁰ which are also bound states giving rise to dips in the conductance characteristics. There are two major differences between magnetic bound states and quasidonor bound states. First, the former occurs regardless of whether the impurity is attractive or repulsive, whereas the latter forms only when the impurity is attractive. Second, the current patterns are very different. In the case of magnetic bound states, the current circulates around the impurity but in the case of quasidonor states, no such circulating current pattern needs to form.¹ Finally, the carrier-concentration profiles are also quite different. In the case of magnetic bound states, the concentration builds up around the impurity but right at the impurity it drops sharply. This is shown in Fig. 8(c). Although the carriers may accumulate around the impurity even in the case of quasidonor state formation (owing to the buildup of evanescent states), the effect is much less pronounced. Finally, Fig. 8(d) shows the electrostatic-potential profile when a magnetic bound state forms.



(a)



(b)

FIG. 6. The electric-field profile inside the structure of Fig. 3 in the absence of any magnetic field. (a) The impurity is attractive; (b) the impurity is repulsive.

B. Spatial distributions for multiple impurities in single-channelled transport: Nonlocal effects

In structures where multiple impurities are present, vortices form in the current pattern as a result of quantum interference between waves reflected from the walls and various impurities.¹ These vortices are a striking aftermath of nonlocality in phase-coherent quantum transport. They are not centered around impurities since they are a consequence of nonlocal effects, and their positions, as well as strengths, depend on the impurity configuration. In Fig. 9(a), we show the current distribution in a sample with two impurities without any magnetic field present. The structure is the same as that in Figs.

3-8 and the Fermi energy is 2.41 meV. There are two vortices in the current pattern. They have opposite circulations and produce antiparallel localized magnetic moments. As we turn on a magnetic field [see Fig. 9(b)], we see immediate quenching of one vortex accompanied by a slight reduction in the strength of the other. This different behavior for the two vortices can be explained by the fact that for one vortex, the Lorentz force tends to produce a circulation in a sense opposite to that of the vortex. Consequently, even a very low field is sufficient to quench this vortex. The resultant change in the quantum interference between the scatterers causes the other vortex to weaken as well, even though it has a circulation pattern in the same direction as that produced by the

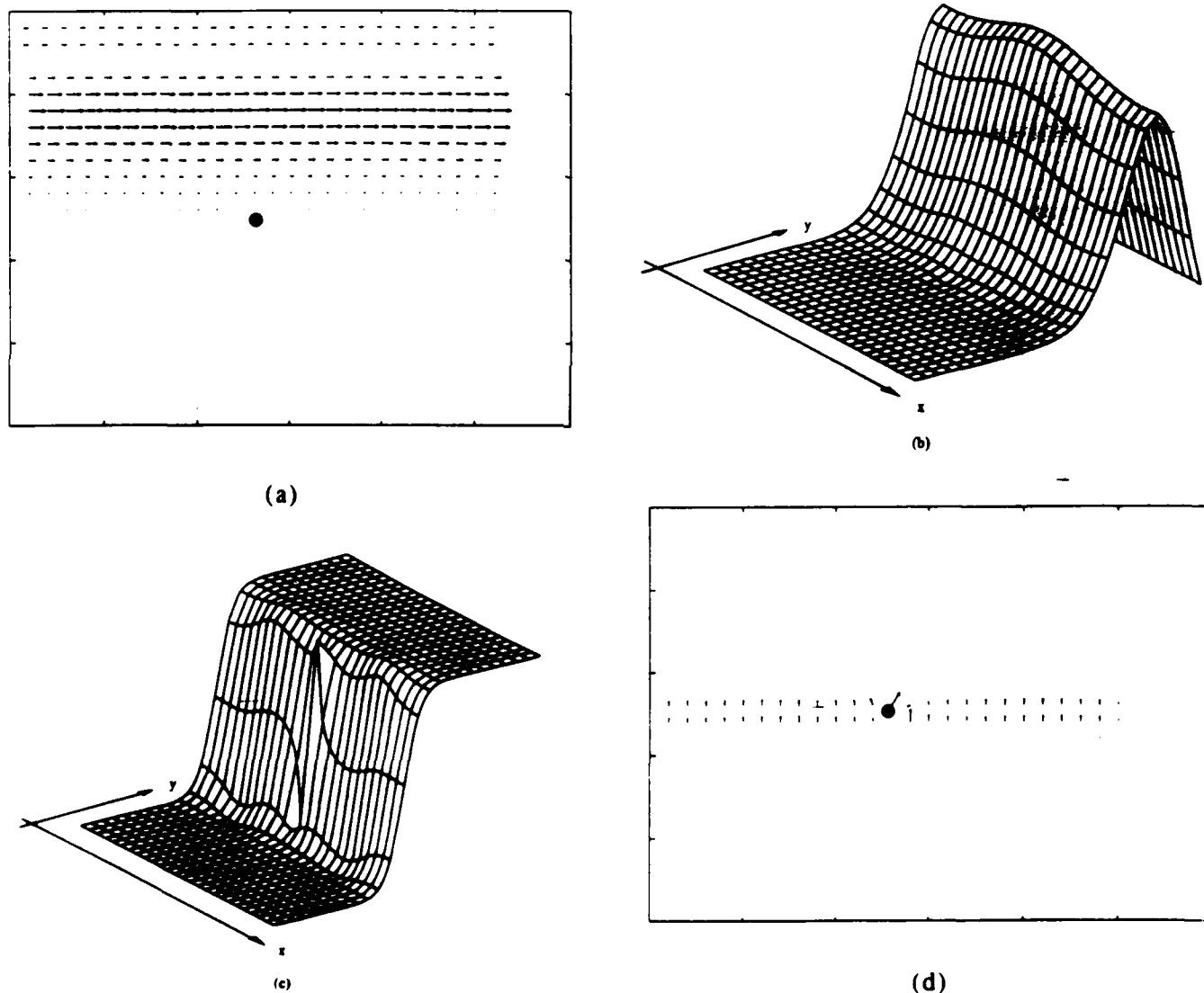


FIG. 7. The spatial distributions in the structure of Fig. 3 in a magnetic-flux density of 2.18 T. The impurity is attractive. (a) The current distribution. The bulk of the forward current is carried by edge states. The small backscattered component near the walls was explained in Ref. 9. This figure should be compared with Fig. 3(a). Note that the local effects of the impurity potential are inhibited by the magnetic field. (b) The Fermi carrier concentration distribution which peaks near an edge because of the Lorentz force skewing the wave function towards that edge. (c) The electrostatic (or chemical) potential profile which shows smooth equipotential surfaces μ_1 along one wall and μ_2 along the other. The longitudinal four-terminal resistance that will be measured by attaching voltage probes near the edges will be zero showing the occurrence of the integer quantum Hall effect. (d) The electric-field distribution which is nonzero only at the center of the wire.

Lorentz force. As the magnetic field is increased slightly, the circulation of one vortex completely reverses while the other weakens further. This is seen in Fig. 9(c). At very high fields, the vortices completely disappear and the current is carried by the edge states from one terminal to the other.

C. Multichanneled transport

The previous results were for single-channelled transport in which an electron can scatter into only one propagating mode (subband). Some of the features observed in the single-channel case may really be consequences of sin-

gle modality (i.e., only intrasubband scattering is allowed as opposed to intersubband scattering). In multichanneled conduction, there are more propagating modes available for scattering so that both intrasubband and intersubband scattering are allowed. This increases the total scattering probability and the current vortices that form require a larger magnetic field to be quenched. There are also more subtle effects. For injection from the second subband, the vortices that form are actually strengthened by a magnetic field. The net magnetic moment of the structure is increased by an external magnetic field so that the response is "paramagnetic." This is in contrast with the situation seen in single-channelled trans-

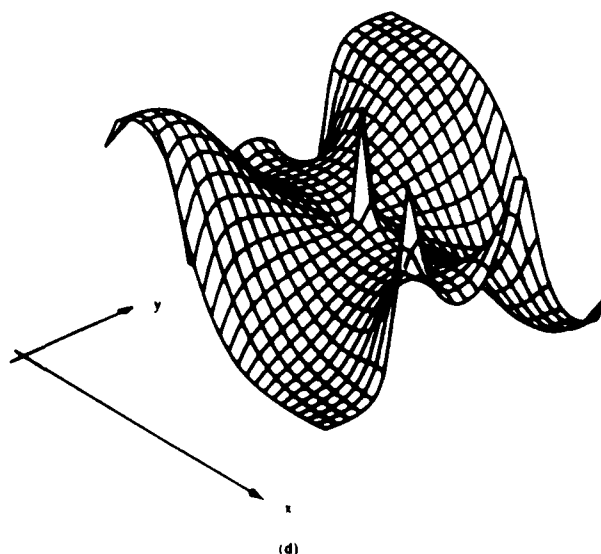
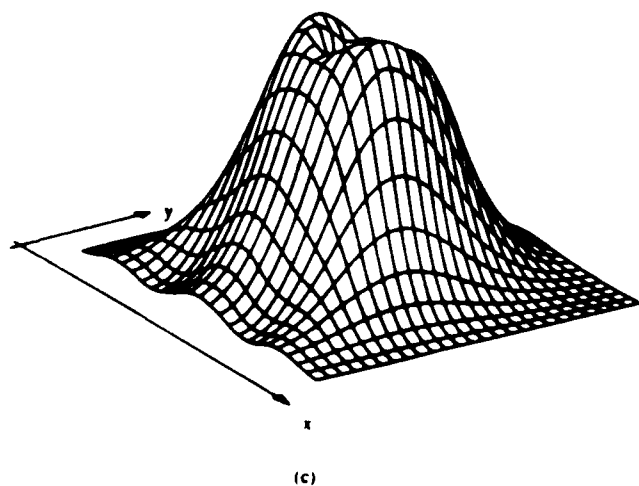
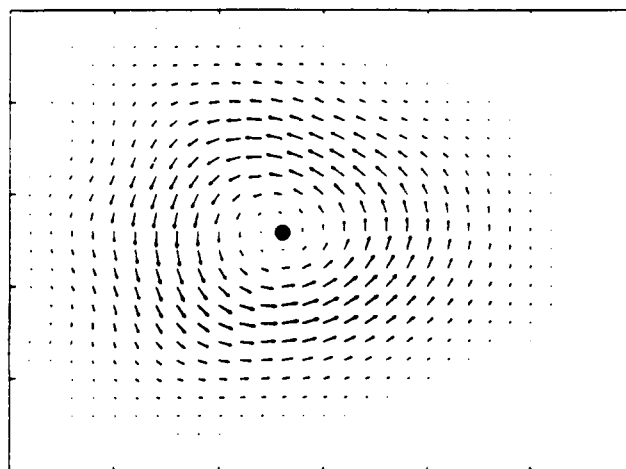
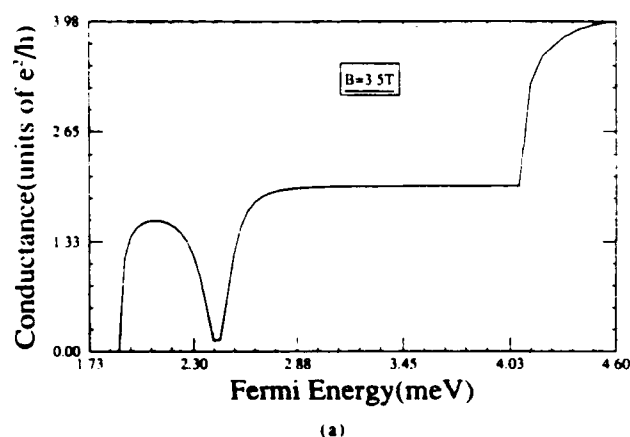


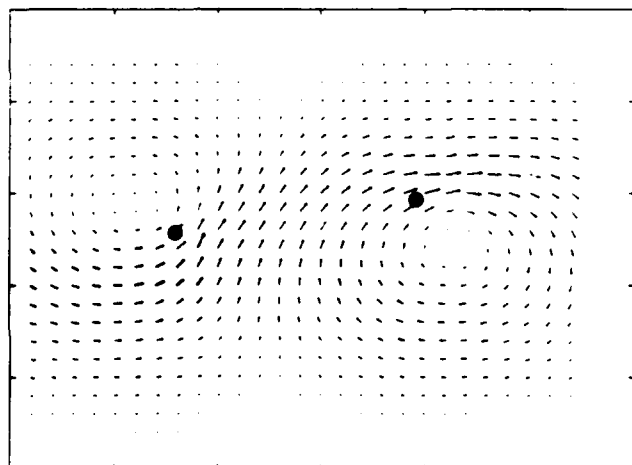
FIG. 8. Results for the formation of a magnetic bound state. (a) The conductance vs Fermi energy for the structure of Fig. 3 containing an attractive impurity. A magnetic-flux density of 3.5 T has been applied. The dip in the conductance occurs at the energy at which a magnetic bound state forms. (b) The current distribution during the formation of a bound state. The current circulates around the impurity. (c) The Fermi carrier concentration distribution. It builds up around the impurity but shows a dip at the impurity location. (d) The electrostatic-potential profile. The equipotential surfaces near the walls are no longer visible [compare with Fig. 7(c)] indicating that the integer quantum Hall effect is destroyed by a magnetic bound state.

port (Figs. 9) in which the response was "diamagnetic."

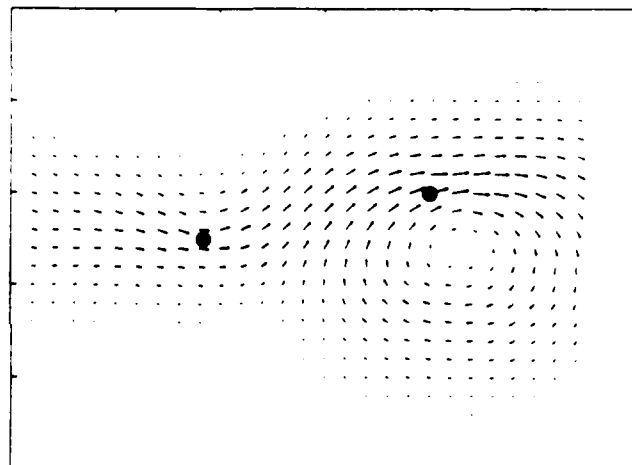
In Fig. 10, we have plotted the conductance versus Fermi energy of a GaAs quantum wire of length 900 Å and width 1000 Å containing four attractive impurities. The conductance shows quantized steps associated with subband filling.¹⁴ The results are plotted for magnetic-flux densities of $B=0$ and 3.5 T. The quantization of the conductance steps is quite poor in the absence of a magnetic field because of significant backscattering from the impurity.^{15,16} When a magnetic field of 3.5 T is turned on, the steps become wider and the lowest steps become much smoother. The steps become wider since the subband separation in energy increases with magnetic field and the widths of the steps are equal to these separations. In addition, the lowest steps become smoother since the transmission probabilities of the lowest subbands increase

at high magnetic fields owing to the formation of edge states with high transmittivity.^{17,18} The quantization for the higher steps is generally worse and does not improve very much with increasing magnetic field since the transmittivities of the higher subbands are always lower. The real-space edge-state trajectories at the two edges corresponding to the higher subbands are spatially closer to one another than those corresponding to the lower subbands. This causes increased coupling between the higher subband edge states and hence increased backscattering.

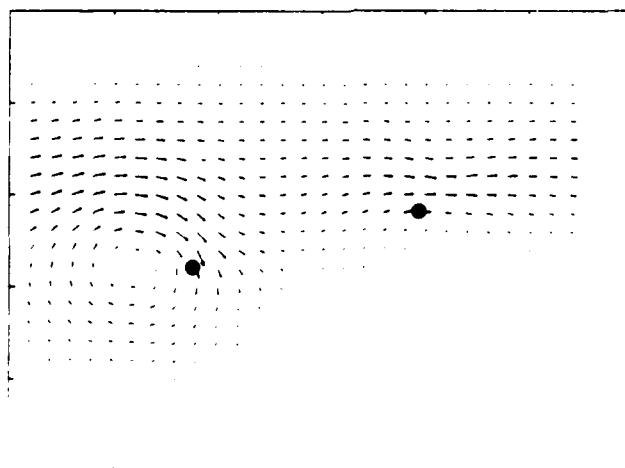
In Figs. 11(a), 11(b), and 11(c), we plot the current distributions around a single attractive impurity when two subbands are occupied. The three figures are the current patterns due to injection from the first subband, the second subband, and the resultant current pattern, re-



(a)



(b)



(c)

FIG. 9. The spatial distributions for a 900-Å-long and 1000-Å-wide structure containing multiple impurities. (a) The current profile in zero magnetic field showing the formation of two strong vortices due to quantum interference between the scatterers and walls. (b) The quenching of one of the vortices and weakening of the other by a magnetic flux density of 0.08 T. (c) Complete change in the vortex patterns and formation of edge states at a flux density of 1 T.

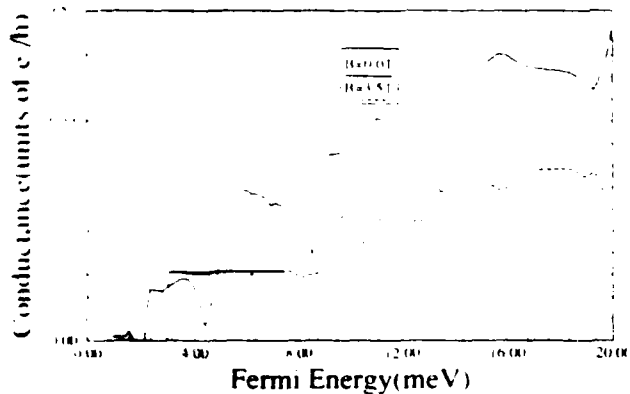


FIG. 10. Recovery of conductance quantization in a disordered structure by a magnetic field. The structure is 900 Å long and 1000 Å wide and contains two attractive impurities. In the absence of a magnetic field, the quantization is quite poor. The transmission of the lowest subband is very small so that the height of the lowest conductance step is also very small. A magnetic-flux density of 3.5 T improves the quantization of the lowest steps significantly and also widens the steps. The quantization of the higher steps is always worse than that of the lower ones because of the relatively lower transmission of the higher subbands.

spectively. The current pattern of the second subband [Fig. 11(b)] does not show any significant backscattering component since the wave function of the second subband has a null at the center of the wire where the impurity is located. Consequently, electrons from the second subband do not interact with the impurity and therefore there is hardly any backscattering. A magnetic field of 1 T skews the wave function of the second subband and causes the electrons from this subband to interact with the impurity. This increases backscattering and hence we expect to see positive magnetoresistance in this case. The current profile for the second subband at a field of 1 T is shown in Fig. 11(d). There is a large backscattered component in the current clearly visible in the upper section of the device which was not present when the magnetic field was absent. Of course, at high enough magnetic fields, edge states form and backscattering is reduced.

D. The integer quantum Hall effect

In Figs. 12(a) and 12(b), we show the electrostatic-potential profiles in a 900-Å-long and 1000-Å-wide quantum wire with two impurities at magnetic fields of 0 and 3.5 T. Only one subband is occupied. We assume that $\mu_1 = 2.413$ meV and $\mu_2 = 2.408$ meV. Figures 12(c) and 12(d) show the corresponding electric-field profiles. Note that in the presence of the magnetic field, both edges of the wire become extraordinarily smooth equipotential surfaces. Consequently, if we attached two voltage probes at any of the edges, they will measure a relative voltage difference of zero and hence the four-terminal linear-response longitudinal resistance will be zero as well. This is a manifestation of the integer quantum Hall

effect. The voltage difference between the two edges is the Hall voltage drop. This figure should be contrasted with Fig. 8(d) which shows the potential profile when a magnetic bound state forms. In that case, the edges are no longer equipotential surfaces and the quantum Hall effect is not observed. It is well known that if the radius of the magnetic bound state is equal to or larger than the width of the wire, then the edges states at the two edges of the wire are coupled by the bound state and electrons can resonantly tunnel from one edge to the other resulting in backscattering. This destroys the quantum Hall effect as predicted by Jain and Kivelson¹⁹ and later verified by Lee, McLennan, and Datta.²⁰ We also observe the same effect in Fig. 8(d).

IV. CONCLUSION

In this paper, we have plotted the current, Fermi carrier concentration, electrostatic-potential, and electric field profiles associated with various magnetotransport phenomena. These results are very helpful in elucidating many features of these effects.

ACKNOWLEDGEMENTS

The authors are indebted to Supriyo Datta, Craig S. Lent, S. M. Goodnick, and Henry Harbury for helpful comments. This work was supported by the U.S. Air Force Office of Scientific Research under Grant No. AFOSR 91-0211 and the Office of Naval Research under Contract No. N00014-91-J-1505.

APPENDIX: DERIVATION OF THE SCATTERING MATRIX FOR A SINGLE IMPURITY IN THE PRESENCE OF A MAGNETIC FIELD

In this Appendix, we derive the scattering matrix describing electron propagation across an elastic δ scatterer (impurity) in the presence of an arbitrary magnetic field. The *real-space matching method* discussed in Ref. 21 is employed. We take a section through the impurity across the width of the structure (y direction) and break this section into a number of mesh points also along the width. Then, we enforce continuity of the wave function and its first derivative across the section.

If we assume the impurity to be located at a position $(0, y)$, then the wave function for the incoming wave from the left side of the impurity can be written as

$$\psi_n^l(x, y) = \zeta_n(x, y) + \sum_{p=1}^M r_{pn} \zeta_{-p}(x, y), \quad x \leq 0.$$

$$\psi_n^r(x, y) = \sum_{p=1}^M t_{pn} \zeta_p(x, y), \quad x \geq 0,$$

where

$$\zeta_{\pm p}(x, y) = \frac{\exp(ik_{\pm p}^F x) \phi_{\pm p}(y)}{\sqrt{|v_p^F|}}.$$

Here r_{ij} and t_{ij} denote the transmission and reflection

coefficients from the j th mode to the i th mode.

For the incoming wave from the right side of the impurity, the wave function can be written as

$$\psi_n^r(x, y) = \zeta_{-n}(x, y) + \sum_{p=1}^M r'_{pn} \zeta_p(x, y), \quad x \geq 0,$$

$$\psi_n^l(x, y) = \sum_{p=1}^M t'_{pn} \zeta_p(x, y), \quad x \leq 0.$$

The scattering matrix across the impurity is defined as

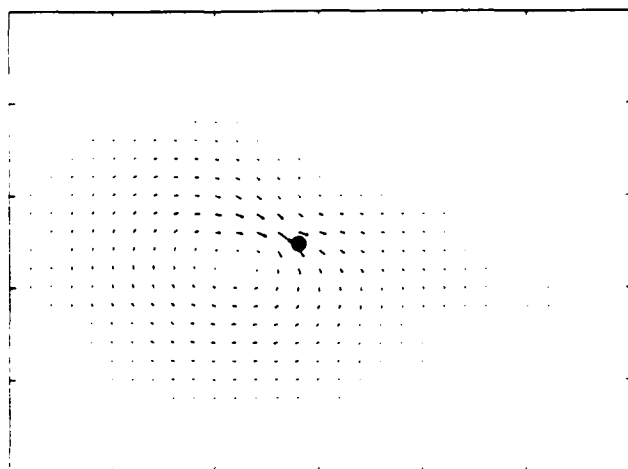
$$S = \begin{bmatrix} t & r' \\ r & t' \end{bmatrix},$$

where the individual elements t, r , etc., are themselves $M \times M$ matrices and M is the total number of modes (propagating plus evanescent).

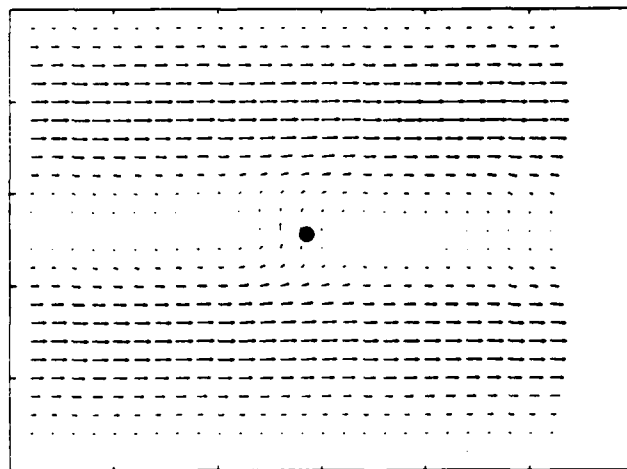
We now enforce the continuity of the wave function and its first derivative at all mesh points. This allows us to write the set of real-space mode-matching equations for the incoming wave from the left

$$\psi_n^l(0, y_i) = \psi_n^r(0, y_i) \quad \text{for all possible } i,$$

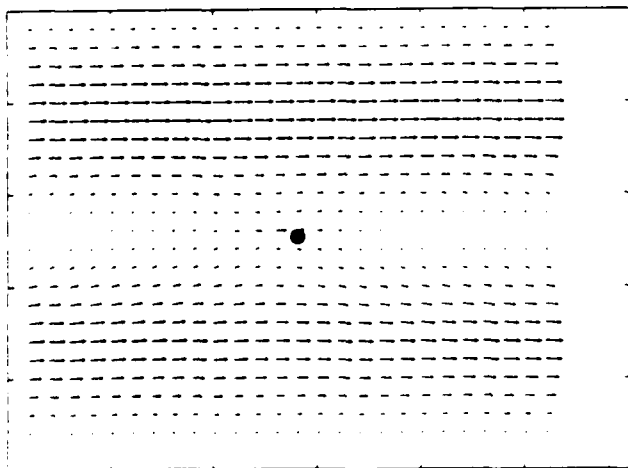
$$\frac{\partial}{\partial x} \psi_n^r(0, y_i) - \frac{\partial}{\partial x} \psi_n^l(0, y_i) = \frac{2m^*}{\hbar^2} \gamma \delta(y - y_0) \psi(0, y_i),$$



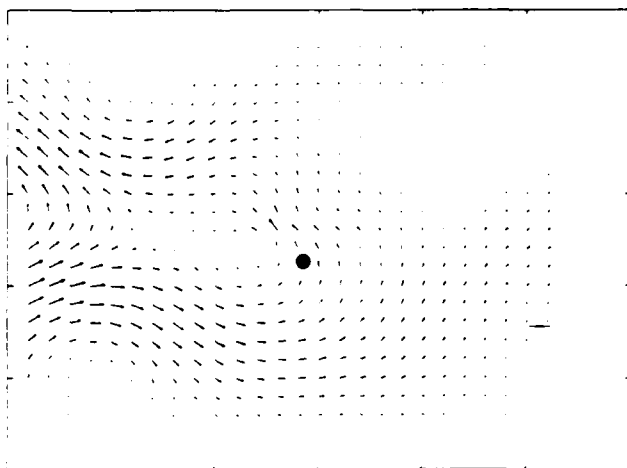
(a)



(b)



(c)



(d)

FIG. 11. Spatial distributions for multichanneled conduction in a structure of length 800 Å and width 1000 Å with one attractive impurity. (a) Current distribution for injection from the lowest subband only at zero magnetic field. (b) Current distribution for injection from the second subband only at zero magnetic field. Not much backscattering is observed since the second subband wave function has a null at the impurity location. (c) The total current distribution at zero magnetic field. (d) The current distribution at a flux density of 1 T for injection from the second subband only. Significant backscattering is observed compared to (b) since the wave function of the second subband is skewed towards the impurity by the magnetic field which increases the interaction of the electrons in the second subband with the impurity.

where y_i is the i th mesh point along the y direction.

At the location of the impurity, the derivative of the wave function is discontinuous by a δ function; elsewhere, it is continuous. Exact treatment of this condition is not possible. We average the quantities over the interval containing the impurity by multiplying every equation by some averaging function and then integrating over the interval. For our case, we simply choose the averaging function as

$$\langle n \rangle = \frac{1}{y_{n+1} - y_n} \int_{y_n}^{y_{n+1}} dy.$$

Substituting for ψ_n and ζ_n , the above equations may be rewritten in matrix notation as

$$\begin{bmatrix} t1 & r1 \\ t2 & r2 \end{bmatrix} \begin{bmatrix} t \\ r \end{bmatrix} = \begin{bmatrix} c \\ d \end{bmatrix},$$

where

$$t1_{ij} = \frac{1}{\sqrt{v_j}} I_{i,j},$$

$$r1_{ij} = -\frac{1}{\sqrt{v_{-j}}} I_{i,-j},$$

$$t2_{ij} = \left[\frac{ik_j}{\sqrt{v_j}} I_{i,j} - \frac{m^* \gamma}{\hbar^2 \sqrt{v_j}} \phi_j(y_0) \delta_{i,0} \right],$$

$$r2_{ij} = \left[\frac{-ik_{-j}}{\sqrt{v_{-j}}} I_{i,-j} - \frac{m^* \gamma}{\hbar^2 \sqrt{v_{-j}}} \phi_{-j}(y_0) \delta_{i,0} \right],$$

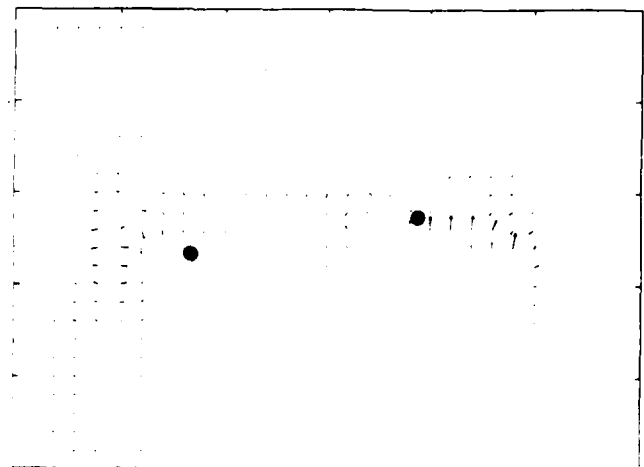
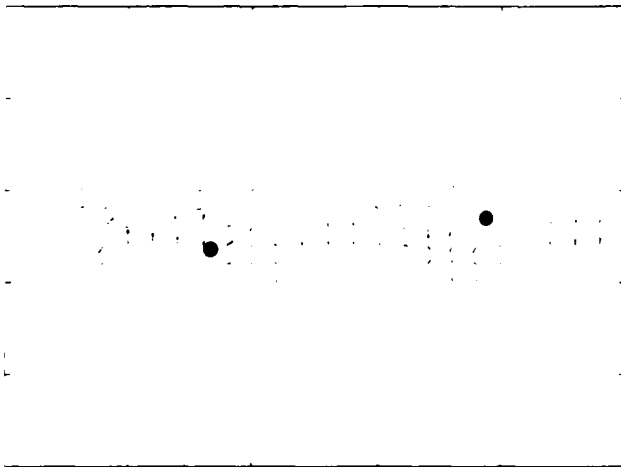
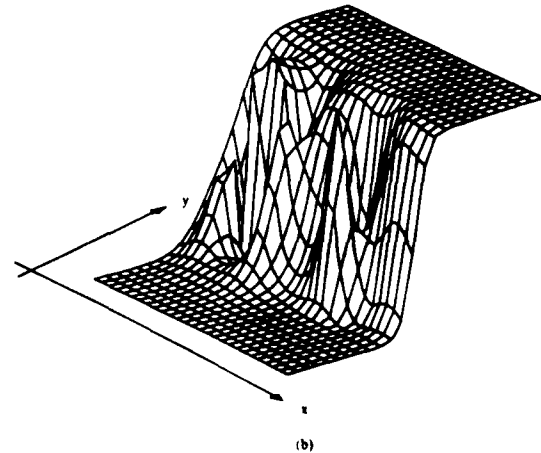
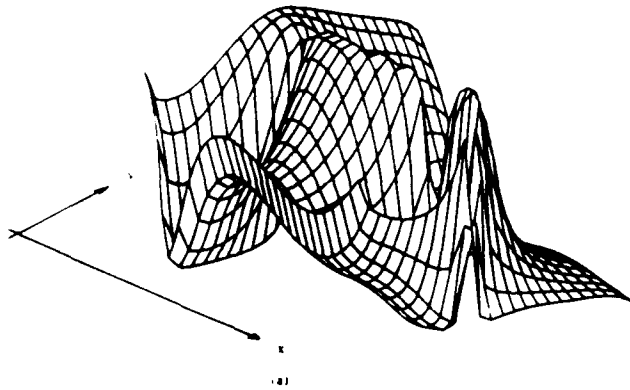
$$c_{ij} = \frac{I_{i,j}}{\sqrt{v_j}}$$

$$d_{ij} = \frac{ik_j}{\sqrt{v_j}} I_{i,j} + \frac{m^* \gamma}{\hbar^2 \sqrt{v_j}} \phi_j(y_0) \delta_{i,0},$$

$$I_{i,j} = \langle i | j \rangle,$$

$$\delta_{i,0} = 1 \quad (y_i = y_0)$$

$$= 0 \quad (y_i \neq y_0).$$



(c)

(d)

FIG. 12. Electrostatic- (or chemical-) potential and electric-field profiles inside a structure with two attractive impurities. (a) Potential profile in the absence of any magnetic field. Significant potential variations are observed. (b) The potential profile at a magnetic-flux density of 3.5 T. The edges have become equipotential surfaces showing the occurrence of the integer quantum Hall effect. (c) The electric-field profile in the absence of any magnetic-field. (d) The electric-field profile at a magnetic-flux density of 3.5 T.

In a similar manner, the equations for the transmission and reflection coefficients for the incoming waves from the right can be derived.

$$\begin{bmatrix} t1' & r1' \\ t2' & r2' \end{bmatrix} \begin{bmatrix} t' \\ r' \end{bmatrix} = \begin{bmatrix} c' \\ d' \end{bmatrix},$$

where

$$t1'_{ij} = \frac{1}{\sqrt{v_{-j}}} I_{i,-j},$$

$$r1'_{ij} = -\frac{1}{\sqrt{v_j}} I_{i,j},$$

$$t2'_{ij} = \left[\frac{ik_{-j}}{\sqrt{v_{-j}}} I_{i,-j} - \frac{m^* \gamma}{\hbar^2 \sqrt{v_{-j}}} \phi_j(y_0) \delta_{i,0} \right],$$

$$r2'_{ij} = \left[\frac{-ik_j}{\sqrt{v_j}} I_{i,j} - \frac{m^* \gamma}{\hbar^2 \sqrt{v_j}} \phi_j(y_0) \delta_{i,0} \right],$$

$$c'_{ij} = \frac{I_{i,j}}{\sqrt{v_j}},$$

$$d'_{ij} = \frac{ik_j}{\sqrt{v_j}} I_{i,j} + \frac{m^* \gamma}{\hbar^2 \sqrt{v_j}} \phi_j(y_0) \delta_{i,0}.$$

For a free propagating region of length x which contains no impurity, we can write down the expressions for transmission and reflection coefficient matrices directly as

$$t_{ij} = \exp(ik_j x) \delta_{ij},$$

$$r_{ij} = r'_{ij} = 0,$$

$$t'_{ij} = \exp(ik_{-j} x) \delta_{ij}.$$

Current conservation requires that the scattering matrix be unitary and we must check for it at all impurity locations. The composite scattering matrix for the whole structure can be found by cascading the individual scattering matrices according to the law of composition of scattering matrices:^{22,23}

$$S = S_1^{\text{free}} \otimes S_1^{\text{imp}} S_2^{\text{free}} \cdots \otimes S_N^{\text{imp}} \otimes S_{N+1}^{\text{free}},$$

where N is the total number of impurities.

A few statements about the numerics may be worthwhile. At high magnetic fields, the coefficient matrix required for finding the values of t and r becomes nearly singular and hence special numerical techniques are needed for matrix inversion. However, we found that the real-space matching technique is always numerically much better behaved than the momentum space mode-matching methods followed in Ref. 24. In fact, the real-space matching method ensures that current continuity is preserved to a much better approximation than momentum space mode-matching methods.

From the set of equations above, it is easy to show that the current across the δ scatterer is conserved and that the scattering matrix is mathematically unitary. To do this, we take the continuity equation and the derivative equations at any point, multiply the right side and the left side together, and then sum them over all points. This immediately shows the current continuity. We mention that the following relationships are useful in deriving the current continuity:

$$\int_{y=0}^{y=W} \phi_j(y) \phi_k(y) \left[k_j^F + k_k^F - \frac{2y}{l^2} \right] dy = 0, \quad j \neq k,$$

$$\int_{y=0}^{y=W} |\phi_j(y)|^2 dy = 1, \quad 1 \leq j \leq N.$$

¹S. Chaudhuri, S. Bandyopadhyay, and M. Cahay, Phys. Rev. B **45**, 11 126 (1992).

²C. S. Lent, Appl. Phys. Lett. **67**, 1678 (1990); Zhen-Li Ji and K. F. Berggren, Phys. Rev. B **45**, 6652 (1992); Zhen-Li Ji, Semicond. Sci. Technol. **7**, 198 (1992).

³S. Chaudhuri and S. Bandyopadhyay, J. Appl. Phys. **71**, 3027 (1992).

⁴For a review, see M. Büttiker, IBM J. Res. Dev. **32**, 317 (1988).

⁵A. M. Kriman, N. C. Kluksdahl, and D. K. Ferry, Phys. Rev. B **36**, 5953 (1987).

⁶C. S. Chu and R. S. Sorbello, Phys. Rev. B **42**, 4928 (1990).

⁷O. Entin-Wohlman, C. Hartzstein, and Y. Imry, Phys. Rev. B **34**, 921 (1986).

⁸M. J. McLennan, Y. Lee, and S. Datta, Phys. Rev. B **43**, 13 846 (1991). This paper presents numerous spatial patterns that are calculated quantum mechanically in the limit of incoherent transport (i.e., in the presence of phase-randomizing scattering).

⁹R. Landauer, IBM J. Res. Dev. **1**, 223 (1957); Z. Phys. B **21**, 247 (1975).

¹⁰P. F. Bagwell, Phys. Rev. B **41**, 10 354 (1990); A. Kumar and P. F. Bagwell, Solid State Commun. **75**, 949 (1990).

¹¹C. S. Lent, Phys. Rev. B **43**, 4179 (1991).

¹²M. Büttiker, Phys. Rev. B **38**, 9375 (1988).

¹³T. Ando and Y. Uemura, J. Phys. Soc. Jpn. **36**, 959 (1974); H. Tamura and T. Ando, Phys. Rev. B **44**, 1792 (1991).

¹⁴B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhoven, D. van der Marel, and C. T. Foxon, Phys. Rev. Lett. **60**, 848 (1988); D. A. Wharam, T. J. Thornton, R. Newbury, M. Pepper, H. Ahmed, J. E. F. Frost, D. G. Hasko, D. C. Peacock, D. A. Ritchie, and G. A. C. Jones, J. Phys. C **21**, L209 (1988).

¹⁵S. He and S. Das Sarma, Phys. Rev. B **40**, 3379 (1989); E. G. Haanapel and D. van der Marel, *ibid.* **39**, 5435 (1989); C. S. Chu and R. S. Sorbello, *ibid.* **40**, 8559 (1989); E. Tekman and S. Ciraci, *ibid.* **42**, 9098 (1990).

¹⁶I. Kander, Y. Imry, and U. Sivan, Phys. Rev. B **41**, 12 941 (1990).

¹⁷B. J. van Wees, L. P. Kouwenhoven, H. van Houten, C. W. J. Beenakker, J. E. Mooij, C. T. Foxon, and J. J. Harris, Phys. Rev. B **38**, 3625 (1988).

¹⁸T. Ando, Phys. Rev. B **44**, 8017 (1991); K. Shepard, *ibid.* **44**, 9088 (1991).

¹⁹J. K. Jain and S. Kivelson, Phys. Rev. Lett. **60**, 1542 (1988).

- ²⁰Y. Lee, M. J. McLennan, and S. Datta, Phys. Rev. B **43**, 14 333 (1991).
- ²¹R. Frohne and S. Datta, J. Appl. Phys. **64**, 4086 (1988).
- ²²M. Cahay, M. McLennan, and S. Datta, Phys. Rev. B **37**, 10 125 (1988); M. Cahay, Ph.D. thesis, Purdue University, 1987.
- ²³R. Redheffer, J. Math. Phys. **41**, 1 (1962).
- ²⁴A. M. Krizan and P. Ruden, Phys. Rev. B **32**, 8013 (1985); D. G. Ravenhall, H. W. Wyld, and R. L. Shult, Phys. Rev. Lett. **62**, 1780 (1989); Y. Takagaki and D. K. Ferry, Phys. Rev. B **44**, 8399 (1991).

Bistable saturation in coupled quantum dots for quantum cellular automata

Craig S. Lent, P. Douglas Tougaw, and Wolfgang Porod

Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

(Received 5 May 1992; accepted for publication 23 November 1992)

A simple model quantum dot cell containing two electrons is analyzed as a candidate for quantum cellular automata implementations. The cell has eigenstates whose charge density is strongly aligned along one of two directions. In the presence of the electrostatic perturbation due to a neighboring cell, the ground state is nearly completely aligned (polarized) in one direction only. The polarization is a highly nonlinear function of the perturbing electrostatic fields and shows the strong bistable saturation important for cellular automation function.

Nanoscale quantum structures with potential device applications have been an active area of exploration for several years. A frequent criticism¹ of many of these structures is the absence of the saturating behavior which forces conventional transistor elements into one of two stable states, "on" or "off." Such bistable saturation is important to keep device performance robust in the presence of physical inhomogeneities and noise.²

The possibility of realizing cellular automata (CA) with regular arrays of quantum dots has been suggested by Bate and others.³ In one example, the necessary nonlinear response of each dot is the result of resonant tunneling through the dot.⁴ We focus on a different paradigm in which each cell of the CA is composed of groups of coupled quantum dots. The confining potentials are such that electrons can tunnel between dots in the same cell but not between different cells. Quantum mechanics and the Coulomb interaction in each cell determine the possible cell states. The Coulomb interaction between electrons in different cells provides a local intercellular coupling mechanism. The nonlinear response of the cell to its electrostatic environment must be a feature of the internal cell dynamics. Recent success in fabricating arrays of very small quantum dots with one or two electrons per dot⁵ prompts us to investigate possible few-electron coupled-dot cell geometries which provides the sort of bistable saturation so desirable. In this letter, we analyze a possible cell geometry with two electrons in the cell. We show that quantum confinement and the intracellular Coulomb interaction together yield the nonlinear saturation behavior which is essential.

We examine a simple nanostructure model cell containing five coupled quantum dots. The model cell is shown schematically in Fig. 1. It consists of a central site and four neighboring sites. Tunneling is possible both between the outer sites and the central site, and between adjacent outer sites. We first consider such a cell holding two electrons (the contrasting case of single and triple cell occupancy is discussed below). We show below that the Coulomb repulsion between the two electrons causes the ground state of the system to be one in which the electrons occupy antipodal sites.

We model the cell using a Hubbard-type Hamiltonian with Coulomb repulsion. The Hamiltonian for a single isolated cell can be written,

$$H_0^{\text{cell}} = \sum_{i,\sigma} E_0 n_{i,\sigma} + \sum_{i,j,\sigma} t_{i,j} (a_{i,\sigma}^\dagger a_{j,\sigma} + a_{j,\sigma}^\dagger a_{i,\sigma}) + \sum_i E_Q n_{i,\uparrow} n_{i,\downarrow} + \sum_{i>j,\sigma,\sigma'} V_Q \frac{n_{i,\sigma} n_{j,\sigma'}}{|r_i - r_j|}, \quad (1)$$

where the number operator $n_{i,\sigma} = a_{i,\sigma}^\dagger a_{i,\sigma}$ and the operator $a_{i,\sigma}^\dagger$ creates an electron at site i with spin σ . The cell parameters which define the Hamiltonian are then the on-site energy, E_0 , the tunneling energies, $t_{i,j}$, and the on-site Coulomb charging energy, E_Q . The parameter V_Q is determined by fundamental constants and the dielectric constant of the material in which the dots are formed. A fixed positive charge $\bar{\rho}$ is assumed at each site sufficient to maintain overall cell charge neutrality. For an isolated cell, this only renormalizes E_0 , but it is important in calculating the interaction between cells as is done below.

For the numerical results we discuss here we choose parameters based on a simple, experimentally accessible model. We consider a cell in a semiconductor with $m^* = 0.067m_0$, which is composed of circular quantum dots of diameter $D = 10$ nm. The near-neighbor distance between the cells is 20 nm. The dielectric constant for the semiconductor is 10. We take $t = 0.3$ meV for coupling to the center site and $t = 0.03$ meV for coupling between outer dots. These tunneling energies can be varied greatly by adjusting the potential barriers between dots. We take $E_Q = V_Q/(D/3)$. We will assume here that the two electrons in the cell have antiparallel spins. The parallel spin case yields results which are qualitatively very similar.

The eigenstates of the Hamiltonian [Eq. (1)] can now be calculated for this specific choice of cell parameters. The Hamiltonian is diagonalized directly in the basis of few-electron states. From the two-electron wave function we calculate the single particle density at each site, ρ_i by find-

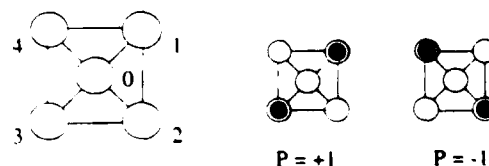


FIG. 1. The quantum cell consisting of five quantum dots which are occupied by two electrons. The mutual Coulomb repulsion between the electrons results in bistability between the $P = +1$ and $P = -1$ states.

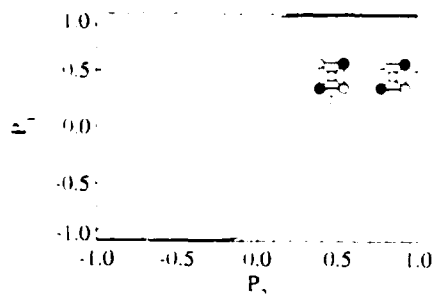


FIG. 2. The cell-cell response function. The induced polarization of cell 1 is shown as a function of the polarization in a neighboring cell 2 (inset). The solid line represents the polarization of the ground state and the dashed line represents the polarization of the first excited state.

ing the expectation value of the total number operator, $n_i = n_{i,\uparrow} + n_{i,\downarrow}$, at each site.

It is helpful to define a scalar quantity which represents the degree to which the electron density is aligned either along the line through sites 1 and 3, or along the line through sites 2 and 4. To this end we define the polarization of the cell as

$$P = \frac{(\rho_1 + \rho_3) - (\rho_2 + \rho_4)}{\rho_0 + \rho_1 + \rho_2 + \rho_3 + \rho_4} \quad (2)$$

If sites 2 and 4 are vacant, the cell is completely in the $P = +1$ polarized state as shown in Fig. 1. If sites 1 and 3 are vacant, the cell is completely in the $P = -1$ polarized state. Clearly if the on-site energies are the same for all sites, the ground state is degenerate, comprising a combination of both polarizations, with no polarization preferred.

We examine the polarization of the low-lying eigenstates of the cell when perturbed by the presence of a nearby cell. We denote the target cell as cell 1 and the perturbing cell as cell 2. The potential at each site i of cell 1 is altered by the Coulomb interaction with the charge $\rho_{2,j}$ at site j of cell 2. The Hamiltonian for cell 1 can be written as the sum of the isolated cell Hamiltonian and a perturbation due to cell 2.

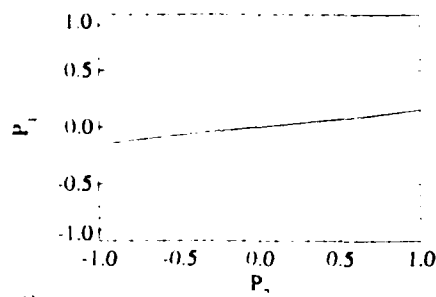
$$H^{\text{cell}} = H_0^{\text{cell}} + H_{12}^{\text{cell}}, \quad (3)$$

where

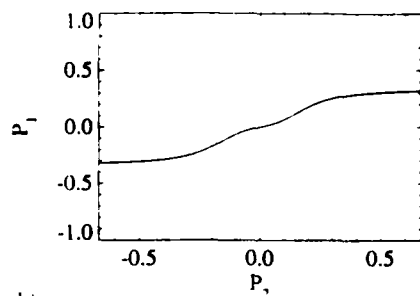
$$H_{12}^{\text{cell}} = \sum_{i,j,\sigma} V_Q \frac{\rho_{2,j} - \bar{\rho}}{|\mathbf{R}_{2,j} - \mathbf{R}_{1,i}|} n_{i,\sigma} \quad (4)$$

Here $\mathbf{R}_{m,i}$ denotes the position of site i in cell m . We solve for the eigenstates of the Hamiltonian [Eq. (3)] as the polarization of cell 2 is varied in the range $P_2 = [-1, 1]$. The occupancy of the central site in cell 2 is assumed to be zero⁶ so that the charge densities, $\rho_{2,j}$, are simple functions of the polarization P_2 . The distance between cell centers is three times the near-neighbor distance in a cell. For each value of P_2 , we find the eigenstates and the associated charge densities and polarizations (Eq. 2). The result is the cell-cell response function—the polarization of cell 1 induced by a polarization of cell 2.

Figure 2 shows the polarization P_1 of the lowest two cell eigenstates as a function of the perturbing cell polar-



a)



b)

FIG. 3. The cell-cell response function for other cell occupancies. (a) A single electron in the cell. (b) Three electrons (two spin up and one spin down) in the cell.

ization P_2 . This figure displays the central result of this letter—that the cell-cell response function is highly nonlinear and bistable. Even a very slight polarization of a cell induces nearly complete polarization of a neighbor.

The strongly nonlinear saturation of the polarization does not occur if only a single electron is in the cell.⁷ Figure 3(a) shows the cell-cell response function for the ground state of a cell occupied by a single electron. The polarization is a very weak and nearly linear function of the perturbation.⁸ The response function for a triply occupied cells (two spin up, one spin down) is shown in Fig. 3(b). Although the response is clearly not linear, it is not nearly as strong as the two-electron case. The bistable saturation present in the doubly occupied cells is a result of the distinct splitting of the degenerate ground state by the perturbation of a neighboring cell.

The rapid saturation of the polarization is the essential nonlinear effect which suggests this type of cell could provide the basis for a CA-type array. In such an array of cells, the differing polarizations of neighboring cells would provide the electrostatic perturbation which leads to a definite polarization of the ground state of the cell. It is possible to extract a CA rule set by finding the cell polarization induced from the various combinations of neighboring polarizations. This process, and the behavior of arrays of quantum cells will be discussed at greater length elsewhere.⁹

It must be noted that to date, quantum dot fabrication techniques have produced dots which tend to be rather far apart, thus only rather weak Coulombic coupling exists.⁴ Our analysis presumes that fabrication techniques will shortly overcome these difficulties, possibly through macromolecular rather than semiconductor implementations.¹⁰

In conclusion, we have shown that strongly nonlinear saturation effects occur in a model two-electron nanoscale system. The charge density "snaps" into one of two positions, depending sensitively on asymmetries in the surrounding charge. This type of very desirable bistable saturation suggests the possibility of quantum cellular automata arrays based on this type of cell.

This work was supported in part by the Air Force Office of Scientific Research, the Office of Naval Research, and by the National Science Foundation under Grant Number ECS890025 through the National Center for Computational Electronics, and utilized the Cray-2 at the National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign. This material is based upon work supported under a National Science Foundation Graduate Fellowship.

¹R. Landauer, *Physics Today* **42**, 119 (1989).

²K. Obermayer, G. Mahler, and H. Haken, *Phys. Rev. Lett.* **58**, 1792 (1987).

R. T. Bate, *Bull. Am. Phys. Soc.* **22**, 407 (1977); J. N. Randall, M. A. Reed, and G. A. Frazier, *J. Vac. Sci. Technol. B* **7**, 1398 (1989); D. K. Ferry, L. A. Akers, and E. W. Greeneich, *Ultra Large Scale Interconnected Microelectronics* (Prentice Hall, Englewood Cliffs, NJ, 1988).

⁴J. N. Randall, A. C. Seabaugh, Y.-C. Kao, J. H. Luscombe, and B. L. Newell, *J. Vac. Sci. Technol. B* **9**, 2893 (1991).

⁵B. Meurer, D. Heitmann, and K. Ploog, *Phys. Rev. Lett.* **68**, 1371 (1992).

⁶A polarization with magnitude near 1 can be achieved only if there is negligible charge density on the central site. That such strong polarization is likely is justified by the following results which show a very strong polarization induced by a neighboring cell.

⁷C. S. Lent, *Nanostructures and Mesoscopic Systems*, edited by W. P. Kirk and M. A. Reed (Academic, Boston, 1992), p. 183.

⁸Single electron cells have been discussed by P. Bakshi, D. A. Broido, and K. Kempa, *J. Appl. Phys.* **70**, 5150 (1991). The lack of the fundamental bistability in the response of a singly occupied cell makes their approach less promising.

⁹C. S. Lent, P. D. Tougaw, and W. Porod (unpublished).

¹⁰For a survey see *Molecular Electronics: Materials and Methods*, edited by P. I. Lazarev (Kluwer, Boston, 1991).

Quantum cellular automata

Craig S Lent, P Douglas Tougaw, Wolfgang Porod and Gary H Bernstein

Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556, USA

Best Available Copy

Received 1 August 1992, accepted for publication 24 December 1992

Abstract. We formulate a new paradigm for computing with cellular automata (CAs) composed of arrays of quantum devices—quantum cellular automata. Computing in such a paradigm is *edge driven*. Input, output, and power are delivered at the edge of the CA array only; no direct flow of information or energy to internal cells is required. Computing in this paradigm is also *computing with the ground state*. The architecture is so designed that the ground-state configuration of the array, subject to boundary conditions determined by the input, yields the computational result. We propose a *specific realization* of these ideas using two-electron cells composed of quantum dots, which is within the reach of current fabrication technology. The charge density in the cell is very highly polarized (aligned) along one of the two cell axes, suggestive of a two-state CA. The polarization of one cell induces a polarization in a neighboring cell through the Coulomb interaction in a very non-linear fashion. Quantum cellular automata can perform useful computing. We show that AND gates, OR gates, and inverters can be constructed and interconnected.

1. Introduction

The continual down-scaling of device dimensions in microelectronics technology has led to faster devices and denser circuit arrays with obvious benefits to chip performance. Dramatic as they have been, these changes have been evolutionary in nature in that even the most advanced chips use the same paradigms for computing as their more primitive ancestors. There is now much expectation that the availability of very dense device arrays might lead to new paradigms for information processing based on locally-interconnected architectures such as cellular automata (CA) and cellular neural networks [1].

There has also been considerable interest in quantum mesoscopic structures for their possible application as devices [2]. Much has been learned about the behavior of electrons flowing through very small structures in semiconductors. Various investigators have pointed out the natural link between mesoscopic quantum systems and cellular automata architectures [3–5]. Because quantum structures are necessarily so small, it is difficult to conceive of a regime in which a single quantum device could drive many other devices in subsequent stages [6]. Furthermore, the capacitance of ultra-small wires forming the connections to each device would tend to dominate the behavior of an assembly of quantum devices. For these reasons locally interconnected structures such as cellular neural networks and CAs may provide the natural architecture for quantum devices.

We focus here on the idea of employing CA architectures which are compatible with nanometer-scale

quantum devices—thus, *quantum cellular automata* (QCA). A QCA would consist of an array of quantum device cells in a locally-interconnected architecture. The cell state becomes identified with the quantum state of the mesoscopic device. Two-state CAs are attractive because they naturally admit to encoding binary information. For a two-state QCA, each cell should have two stable quantum states. The state of a given cell should influence the state of the neighboring cells. Two ingredients are essential then: (1) the bistability of the cell, and (2) coupling to neighboring cells.

We propose a cell which is composed of coupled quantum dots occupied by two electrons [7]. The requisite bistability is accomplished through the interaction of quantum confinement effects, the Coulomb interaction between the two electrons, and the quantization of charge [8]. The intercellular interaction is provided by the Coulomb repulsion between electrons in different cells. We analyze this cell and the interactions between neighboring cells in section 2.

In section 3 we propose a new paradigm for how computation could be performed with an array of quantum devices. Because no direct connections can be made to interior cells, information or energy can enter the array only from the edges. *Edge-driven computation* imposes further constraints on the nature of the computing process [9]. The lack of direct connections to the interior cells also means that no mechanism exists for keeping the array away from its equilibrium ground-state configuration. We are therefore led to use the ground state of the array to do the computation. *Computing with the ground state* means that the physics of the array must

perform the computation by dissipating energy as it relaxes to the ground state. This has the distinct advantage that the computing process is independent of the details of the energy relaxation mechanisms and that the unavoidable energy dissipation is useful to the computing process.

Section 4 demonstrates that QCAS can perform useful functions. We show how logical gates and inverters can be constructed with arrays of the two-electron bistable quantum cell we propose. Section 5 discusses some key issues in realizing QCAS as a viable technology and section 6 identifies technological advantages that a successful QCA implementation would enjoy.

2. Few-electron quantum cells

The specific cell we consider here is shown in figure 1. Four quantum dots are coupled to a central dot by tunnel barriers. The two electrons tend to occupy antipodal sites in one of two configurations, shown in the figure as the $P = +1$ and $P = -1$ configurations. Our analysis below will show that the cell is indeed in one of these two stable states, and that an electrostatic perturbation, perhaps caused by neighboring cells, switches the cell between these two states in a very abrupt and non-linear way. This permits the encoding of bit information in the cell.

The essential ingredients that produce the bistable saturation behavior [10] which is so desirable are (1) quantum confinement, (2) Coulomb interaction between electrons, (3) few-electron quantum mechanics, and (4) the discreteness of electronic charge.

2.1. A model for the quantum cell

We model the cell shown in figure 1 using a Hubbard-type Hamiltonian. For the isolated cell, the Hamiltonian can be written

$$H_0^{\text{cell}} = \sum_{i,\sigma} E_{0,i} n_{i,\sigma} + \sum_{i,\sigma} t(a_{i,\sigma}^\dagger a_{0,\sigma} + a_{0,\sigma}^\dagger a_{i,\sigma}) + \sum_i E_Q n_{i,\uparrow} n_{i,\downarrow} + \sum_{i>j,\sigma,\sigma'} V_Q \frac{n_{i,\sigma} n_{j,\sigma'}}{|R_i - R_j|} \quad (1)$$

Here $a_{i,\sigma}$ is the annihilation operator which destroys a particle at site i ($i = 0, 1, 2, 3, 4$) with spin σ . The number operator for site i and spin σ is represented by $n_{i,\sigma}$. The

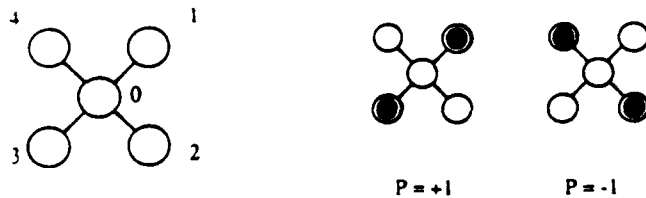


Figure 1. The quantum cell consisting of five quantum dots which are occupied by two electrons. The mutual Coulombic repulsion between the electrons results in bistability between the $P = +1$ and $P = -1$ states.

on-site energy for the i th dot is $E_{0,i}$; the coupling to the central dot is t ; the charging energy for a single dot is E_Q . The last term represents the Coulombic potential energy for two electrons located at sites i and j at positions R_i and R_j . Unless otherwise noted, we will consider the case where all the on-site energies are equal, $E_{0,i} = E_0$.

For our standard model cell, on which the numerical results reported here are based, we obtain the values of the parameters in the Hamiltonian from a simple, experimentally reasonable model. We take each site to be a circular quantum dot with diameter $D = 10$ nm, and take E_0 to be the ground-state energy of such a dot holding an electron with effective mass $m^* = 0.067 m_0$. The near-neighbor distance between dot centers, a , is taken to be 20 nm. The Coulomb coupling strength, V_Q , is calculated for a material with a dielectric constant of 10. We take $E_Q = V_Q/(D/3)$ and $t = 0.3$ meV.

It is useful to define a quantity which represents the degree to which the charge density for a given eigenstate of the system is aligned linearly. This alignment could be either along the line through sites 1 and 3 or along the line through sites 2 and 4. For each site, we calculate the single particle density ρ_i , which is simply the expectation value of the total number operator for the two-electron eigenstate. The polarization, P , is defined as

$$P = \frac{(\rho_1 + \rho_3) - (\rho_2 + \rho_4)}{\rho_0 + \rho_1 + \rho_2 + \rho_3 + \rho_4} \quad (2)$$

For an isolated cell with all on-site energies equal, no polarization is preferred. We will see below that perturbations due to charges in neighboring cells can result in a strongly polarized ground state. The polarization thus defined is not to be confused with the usual dipole polarization of a continuous medium. It simply represents the degree to which the electrons in the cell are aligned and in which of the two possible directions the alignment occurs. For the states of interest here, the cell has no dipole moment.

The interaction of the cell with the surrounding environment, including other neighboring cells, is contained in a second term in the Hamiltonian which we write as $H_{\text{inter}}^{\text{cell}}$. We solve the time independent Schrödinger equation for the state of the cell, $|\Psi_n\rangle$, under the influence of the neighboring cells:

$$(H_0^{\text{cell}} + H_{\text{inter}}^{\text{cell}})|\Psi_n\rangle = E_n|\Psi_n\rangle \quad (3)$$

The spins of the two electrons can be either aligned or anti-aligned, with corresponding changes in the spatial part of the wavefunction due to the Pauli principle. We will restrict our attention to the case of anti-aligned spins here because that is the ground-state configuration; the spin-aligned case exhibits nearly identical behavior. The Hamiltonian is diagonalized directly in the basis of few-electron states. We calculate single-particle densities, ρ_i , from the two-particle ground-state wavefunction $|\Psi_0\rangle$,

$$\rho_i = \sum_{\sigma} \langle \Psi_0 | n_{i,\sigma} | \Psi_0 \rangle \quad (4)$$

and from the densities calculate the resultant polarization P , equation (2). To maintain charge neutrality, a

fixed positive charge, $+1$, with magnitude $1/2$ is assumed at each site. For the isolated cell, this has no effect and is included on the on-site energies. For several cells in close proximity, as will be considered below, the maintenance of overall cell charge neutrality means that the inter-cellular interaction is due to dipole, quadrupole, and higher moments of the cell charge distribution. If cells had a net total charge then electrons in cells at the periphery of a group of cells would tend to respond mostly to the net charge of the other cells.

2.2. The cell-cell response function

To be of use in a CA architecture, the polarization of one cell must be strongly coupled to the polarization of neighboring cells. Consider the case of two nearby cells shown in the inset to figure 2. Suppose we fix the charge distribution in the right cell, labeled cell 2. We assume cell 2 has polarization P_2 , and that the charge density on site 0 is negligible (this means the charge density is completely determined by the polarization). For a given polarization of cell 2, we can compute the electrostatic potential at each site in cell 1. This additional potential energy is then included in the total cell Hamiltonian. Thus the perturbing Hamiltonian is

$$H_{\text{inter}}^{\text{cell}} = H_1^{\text{cell}} = \sum_{i \in \text{cell } 1, \sigma} V_i^1 n_{i, \sigma} \quad (5)$$

where

$$V_i^m = \sum_{k \neq m, j} V_0 \frac{(\rho_j^k - \bar{\rho})}{|\mathbf{R}_{k,j} - \mathbf{R}_{m,i}|} \quad (6)$$

is the potential at site i in cell m due to the charges in all other cells k . We denote the position of site j in cell k as $\mathbf{R}_{k,j}$. The total Hamiltonian for cell 1 is then

$$H^{\text{cell}} = H_0^{\text{cell}} + H_1^{\text{cell}}. \quad (7)$$

The two-electron Schrödinger equation is solved using this Hamiltonian for various values of P_2 . The ground-state polarization of cell 1, P_1 , is then computed as described in the previous section.

Figure 2(b) shows the lowest four eigen-energies of cell 1 as a function of P_2 . The perturbation rapidly separates states of opposite polarization. The excitation energy for a completely polarized cell to an excited state of opposite polarization is about 0.8 meV for our standard cell. This corresponds to a temperature of about 9 K. Figure 2(a) shows P_1 as a function of P_2 —the cell-cell response function. A very small polarization in cell 2 causes cell 1 to be very strongly polarized. This non-linear response is the basis of the OCAs we describe here. As the figure shows, the polarization saturates very quickly. This observation yields two important results:

(i) *The bipolar saturation means that we can encode bit information using the cell polarization.* A cell is almost always in a highly polarized state with $P \approx \pm 1$. We define the $P = +1$ state as a bit value of 1 and the $P = -1$ state as a bit value of 0. Only if the electrostatic environment due to other cells is nearly perfectly sym-

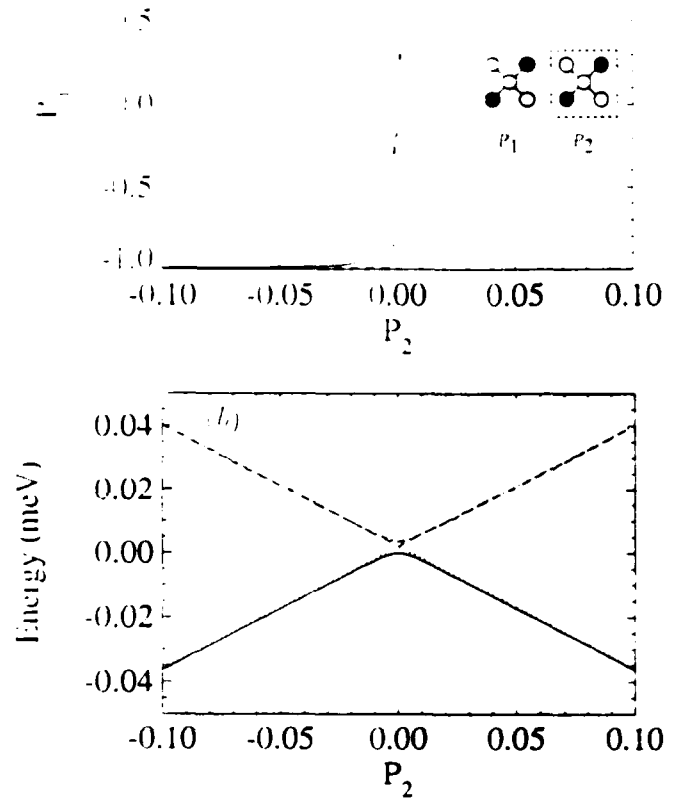


Figure 2. The cell-cell response function. The polarization of the right cell is fixed and the induced polarization in the left cell is calculated. (a) The calculated polarization of cell 1 as a function of the polarization of cell 2. Note that the range of P_2 shown is only from -0.1 to $+0.1$. This is because the transition in the induced polarization is so abrupt. (b) The first four eigen-energies of cell 1. The polarization of the lowest two are shown in (a).

metric will there be no polarization.

(ii) *The polarization of one cell induces a polarization in its neighbor.* Figure 2 shows that even a very slight polarization will induce nearly complete polarization of a neighboring cell. This cell-cell Coulomb coupling provides the mechanism for CA-like behavior. The rapid saturation of the cell-cell response function is analogous to the gain necessary to preserve digital logic levels from stage to stage.

The abruptness of the cell-cell response function depends on the ratio of the dot-to-dot coupling energy, t in equation (1), to the Coulomb energy for electrons on different sites. The magnitude of the coupling depends exponentially on both the distance between the dots and the height of the potential barrier between them [11], each of which can be adjusted as engineering parameters. Figure 3 shows how the cell-cell response function varies with t .

2.3. Self-consistent analysis of several quantum cells

In the analysis of the previous two subsections, the two-electron eigenstates were calculated for a single cell. It is

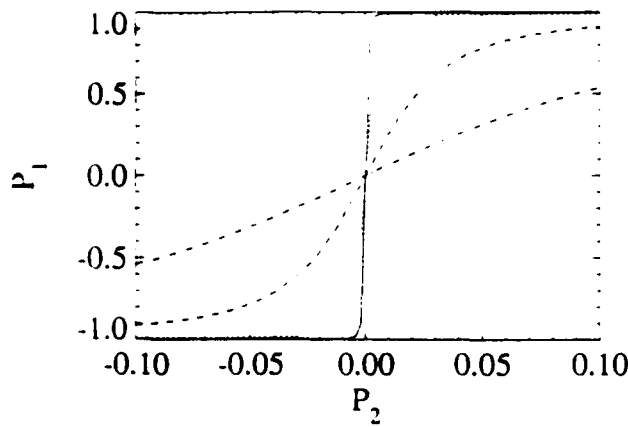


Figure 3. The cell-cell response function for various values of the dot-to-dot coupling energy (t in equation (1)). The induced cell polarization P_1 is plotted as a function of the neighboring cell polarization P_2 . The results are shown for values of the coupling energy, $t = -0.2$ (full curve), -0.3 (dotted curve), -0.5 (dashed curve), and -0.7 (dot-dashed curve) meV. Note that the response is shown only for P_2 in the range $[-0.1, +0.1]$.

important to note that for the Hamiltonian employed, these are exact two-particle eigenstates. Exchange and correlation effects have been included exactly. This was possible because we could explicitly enumerate all possible two-electron states and diagonalize the Hamiltonian in this basis set. We want to analyze clusters and arrays of cells to investigate possible device architectures. To do so we need to calculate the ground-state wavefunction of a group of cells. Exact diagonalization methods are then no longer tractable because the number of possible many-electron states increases so rapidly as the number of electrons increases. We must therefore turn to an approximate technique.

The potential at each site of a given cell depends on the charge density at each site of all other cells. We will treat the charge in all other cells as the generator of a Hartree-type potential and solve iteratively for the self-consistent solution in all cells. This approximation, which we call the intercellular Hartree approximation (ICHA), can be stated formally as follows. Let Ψ_0^k be the two-electron ground-state wavefunction for cell k , and ρ_j^m be the single-particle density at site j in cell m . We begin with an initial guess for the densities. Then, for each cell we calculate the potential due to charges in all other cells.

$$V_i^k = \sum_{m \neq k, j} V_Q \frac{(\rho_j^m - \bar{\rho})}{|R_{m,j} - R_{k,i}|} \quad (8)$$

Although the neighboring cells will normally dominate this sum, we do not examine only near-neighbors but include the effect of all other cells. For each cell k , this results in a perturbation of the basic cell Hamiltonian of equation (1):

$$H_k^{\text{cell}} = \sum_{i \in \text{cell } k, \sigma} V_i^k n_{i,\sigma} \quad (9)$$

The Schrödinger equation for each cell is now solved for

the two-electron ground-state eigenfunction:

$$(H_0^{\text{cell}} + H_k^{\text{cell}})|\Psi_0^k\rangle = E_0^k|\Psi_0^k\rangle. \quad (10)$$

From the ground-state eigenfunctions we calculate the improved single-particle densities.

$$\rho_j^k = \sum_{\sigma} \langle \Psi_0^k | n_{j,\sigma} | \Psi_0^k \rangle. \quad (11)$$

The improved densities are then used in equation (8) and the system is iterated until convergence is achieved. Once the system converges, the many-electron energy, E_{total} is computed from the sum of the cell eigenenergies using the usual Hartree correction term to account for overcounting of the Coulomb interaction energy between cells:

$$E_{\text{total}} = \sum_k E_k^0 - \sum_{k > q, i, j} V_Q \frac{\rho_i^k \rho_j^q}{|R_{k,i} - R_{q,j}|}. \quad (12)$$

It should be stressed that the ICHA still treats Coulombic, exchange, and correlation effects between electrons in the *same* cell exactly. The Hartree mean field approach is used to treat self-consistently the interaction between electrons in *different* cells. Since electrons in different cells are physically distinguishable (there being no wavefunction overlap), the exchange coupling between them is zero. The Hartree and Hartree-Fock approximations are therefore equivalent in this case.

The converged ICHA solution will be an (approximate) eigenstate of the entire system. In general, however, it need not be the ground state. As with the usual Hartree approximation, which of the eigenstates the scheme converges to is determined by the choice of the initial guess. To find the ground state we must try many initial state guesses and determine which converged solution has the lowest energy. Typically, this does not present a serious problem for the type of cellular arrays considered here because the set of likely ground states is easily discerned. In general, a systematic search may be required.

The procedure described above uses, at each stage of the iteration, only the ground-state wavefunction of each cell. If all the excited states of the entire system were desired, we would have to include states composed of excited cell states as well. Since our interest is in the ground state, this is not necessary. It is relevant to point out however, that because each cell is in a 'local' ground state, we do not require coherence of the many-electron wavefunction across the whole array of cells. All that is required to support this analysis is that the wavefunction is *coherent across a single cell*. No information about the phase of the wavefunction in other cells is relevant to the wavefunction in a given cell—only the charge densities in other cells need be known.

3. Computing with quantum cellular automata

We present a new paradigm for computing with QCAs. This represents a complete picture of how quantum devices could be coupled in a CA architecture to perform

useful functions. The paradigm we propose is shown schematically in figure 4. We will focus on the zero-temperature case; temperature effects will be considered below. As shown in the figure, the inputs are along an edge of the array. Specifying the inputs consists of electrostatically fixing the polarization of the input cells. This could be accomplished by simply applying voltages to conducting 'set' lines which come in close proximity to the cells, but any method that fixes the cell polarization state would do. The output cells are not fixed; their polarization state is sensed, perhaps by electrostatic coupling to 'sense' lines. There could also be several input and output edges. Computation proceeds in the following steps:

- (i) Write the input bits by fixing the polarization state of cells along the input edge (edge-driven computation).
- (ii) Allow the array to relax to its ground state with these inputs (computing with the ground state).
- (iii) Read the results of the computation by sensing the polarization state of cells at the output edge.

The essential elements that define this computing paradigm are *computing with the ground state* and *edge-driven computation*, which we discuss below.

3.1. Computing with the ground state

The advantage of computing with the ground state is that it leaves the computing process insensitive to the details of the dissipative processes which couple the electrons in the array to the environment. Consider a QCA at zero temperature for which all the input cells have been held in a fixed state. Dissipative processes have brought the array to its ground-state configuration for these boundary conditions. Suppose at time $t = 0$ the input cell states are set to their new input values completely abruptly. Just after the inputs are applied at the edge of the QCA, the

array is no longer in the ground state but is now in an excited non-stationary state for the new boundary conditions. In the time between 0 and t_r , a characteristic relaxation time, various dissipative processes will bring the array to its new ground-state configuration. After that, the array will be stable until the boundary conditions are changed again. During the relaxation time the temporal evolution of the system is *very* complicated. Even without dissipation, the system will undergo quantum oscillations due to interference between the various eigenstates which compose the $t = 0^+$ state. The dissipative processes, like phonon emission, introduce extraordinary complication in the temporal evolution. The exact state of the system at a particular time $t < t_r$ depends not just on phonon emission rates, but on the *particular* phonons emitted by these *particular* electrons. In short, the temporal evolution before $t = t_r$ depends on the *precise* microscopic details of the dissipative dynamics. By contrast, the ground state configuration to which the system relaxes is *completely independent of the dissipation mechanisms*. Hence we choose to use the ground state only for computing.

3.2. Edge-driven computation

In the QCA computing paradigm we are proposing, the input data is represented by edge cells whose polarization is fixed. Computing then proceeds by allowing the physics interior to the QCA to 'solve' the dissipative many-electron problem for this new set of boundary conditions. The array is designed so the part of the ground-state 'solution' of the many-body problem which appears at the output edge corresponds to the solution of the computing problem posed by the input data.

The advantage of writing input and reading output only at the edges of the array is that no separate connections to the array interior need be made. Because quantum devices are of necessity extremely small, the problem of making contacts to each element or device becomes severe. If a single array contains thousands of individual cells, the 'wiring' problem is overwhelming.

Edge-driven computation is, in fact, the practical requirement which makes computing with the ground state necessary. If no connections can be made to the interior of the array, there is no controlled mechanism for keeping the system away from the ground state. Neither clocking nor refresh mechanisms are available. With a change in input, the system will dissipate energy and find a new equilibrium ground state. The only choice is whether to try to perform computation with the system's transient response, or with its ground state. For the reasons discussed above, the ground-state approach is preferable.

Conventional computing, by contrast, is done using very highly-excited, non-equilibrium states. Because each element (device) can be separately contacted, energy can be fed into the system at each point. The entire system can thereby be maintained in non-equilibrium states. The advantage of this is that the energy difference between the states used for computing can be very much larger than

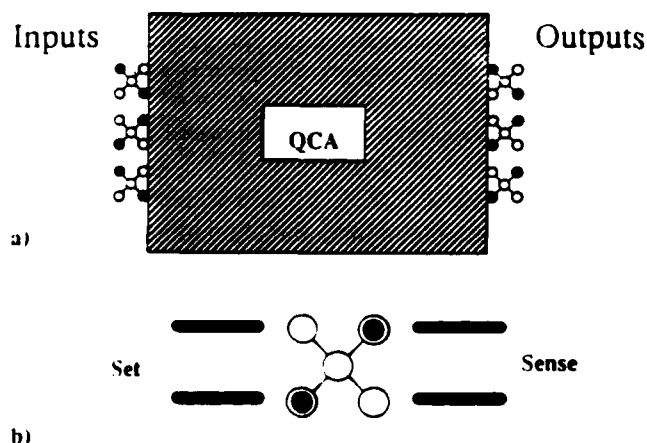


Figure 4. The new paradigm for computing with quantum cellular automata (QCAs). The input to the QCA is provided at an edge by setting the polarization state of the edge cells (*edge-driven computation*). The QCA is allowed to dissipatively move to its new ground-state configuration and the output is sensed at the other edge (*computing with the ground state*). The 'set' and 'sense' lines are shown schematically.

$k_B T$. The requirement that each element be driven far from equilibrium ultimately contributes to the difficulty of reducing the scale of conventional technology to the nanometer level. The breakdown of the operating device physics at small scales also plays a crucial role in the scale-down problem.

Ultimately, temperature effects are the principal problem to be overcome in physically realizing the QCA computing paradigm. The critical energy is the energy difference between the ground state and the first excited state of the array. If this is sufficiently large compared with $k_B T$, the system will be reliably in the ground state after time t_r . Fortunately, this energy difference increases quadratically as the cell dimensions shrink. If the cell size could be made a few Ångströms, the energy differences would be comparable to atomic energy levels—several electron Volts! This is, of course, not feasible with semiconductor implementations, but may ultimately be attainable in molecular electronics. It may, however, be possible to fabricate cells in semiconductors small enough to work reliably at reasonable cryogenic temperatures.

3.3. Relation to synchronous CA rules

The relationship between the QCAs described here and traditional rule-based CAs is not direct. CAs are usually described by a set of CA rules which govern the temporal evolution of the array [12]. Time proceeds in discrete increments called generations. The rules determine the state of the array based on its configuration in the previous generation. Clearly, for the QCA described here, the temporal evolution proceeds not through discrete generations but through continuous physical time. Moreover, as argued above, we are not particularly interested in the temporal evolution of the QCA in order to perform computations. We are only concerned with the final ground-state configuration associated with a particular input state. Like the rule-based synchronous CA, the QCA is an array of interacting multi-state cells and the behavior is dominated by near-neighbor interactions between cells. Thus, the QCA is chiefly related to traditional CAs by *analogy*.

Nevertheless, it is possible to construct a rule-based CA from the QCA interacting cell Hamiltonian (10). The CA so constructed may be useful, perhaps not in describing the transient state of the QCA, but rather in calculating the ground-state configuration, which is our primary concern anyway.

3.4. CA rules from the Schrödinger equation

The CA rule set is constructed as follows. For each cell, consider all possible polarization states ($P = \pm 1$) of the neighbors (neighbors out to any distance useful can be considered). For each configuration of the neighboring polarization, solve the Schrödinger equation (10) and determine the target-cell ground state and its polarization. The map of neighbor polarizations to target-cell polarization constitutes the CA rule set for that particular

target cell. In general, a different rule set may apply to each cell. Typically, many cells will have similar environments and use the same rules.

The rule set obtained by this procedure can be recast in terms of a *weighted-voting* procedure. In deciding the state of a particular cell, the neighboring cells vote according to their own state. The votes are weighted differently depending on the geometrical relationship between each neighbor and the target cell. The votes of closer cells are weighted more heavily than those of more distant cells. In addition, the weights can be negative, indicating that the energetics of the interaction between the neighbor and the target cell favor them having opposite polarizations. The CA rules generated by the solution of the Schrödinger equation for the target cell can then be recast in the form of voting weights for the neighbors. Any set of voting weights which reproduces the CA rule set is equivalent.

3.5. Extended CA rules

This procedure so far has one problem which can be remedied by expanding the rules slightly. It is possible for the votes of the neighbors to result in a 'tie'. That is, the neighboring polarizations may be arranged so symmetrically that the ground-state polarization of the target cell is zero. It is desirable to break this tie by consulting the immediate history of the neighbors. The neighbors which flipped their polarization in the preceding generation are simply weighted more heavily than those which have not flipped. This introduces a notion of momentum which is otherwise absent in a two-state CA. With these *momentum rules*, ties are still possible but are now exceedingly rare events that can be handled by tie-breaking with a random number.

The CA rules corresponding to a particular QCA are thus derived from the Schrödinger equation and augmented by the momentum rule discussed above. The evolution of the synchronous CA is still not directly related to the temporal evolution of the physical QCA. The CA rules know nothing of the details of the dissipative dynamics, for example. However, in our experience, the synchronous CA with the momentum rules can be useful in determining the ground state of the QCA. If we start with a stable QCA state, and then flip the input cells to correspond to the new input condition, the synchronous CA will evolve to a stationary state which corresponds to the ground state of the physical QCA. That the final state is really the ground state can be checked by using the more rigorous self-consistent calculations described in the previous section.

4. Device applications

Two types of QCA structure for computing can be envisioned. One type is a very large regular array of cells. We have work in progress exploring this type of array. It is widely appreciated that computing with large regular CAs is a significant challenge, particularly with a simple

rule set. The solution to this difficult problem may have the greatest long-term potential, however, for exploiting the massive parallelism inherent in the QCA paradigm.

A second type of QCA structure involves a highly irregular array of cells. We show below that using simple irregular arrays one can produce structures analogous to wires, inverters, AND gates and OR gates. Since these can be connected together, more complex devices such as adders and multipliers can be constructed. Because the individual devices are so small, this represents a potentially enormous increase in functional density in an architecture free of the usual interconnect problems. We examine below how these basic logical gates can be constructed from quantum cells.

The device configurations shown are the results of self-consistent calculations of the ground state using the ICHA described above. Figure 5 shows the calculated ground-state charge density on each site of the cellular array. In these figures the dot diameters reflect the relative electron density at each site (dot) in the cell.

4.1. Wires

A linear chain of cells oriented as shown in figure 5(a) functions as a wire, transmitting a 0 or 1 ($P = +1$ or $P = -1$) from one end of the wire to the other. This is demonstrated by fixing the polarization of one end (the left in the figure), while letting the other end be unconstrained, and calculating the self-consistent ground state of the chain using the ICHA method. Figure 5(a) shows the results of this calculation. Not surprisingly, the ground state consists of all cells aligned with the same polarization as the end cell. The first excited state of the chain has a 'kink' in it at the chain center, i.e., half the cells polarized one way and half polarized the other. For our example, the energy of the first excited state is about 1 meV ($\Delta E/k_B T = 10$ K) above the ground-state energy. Wire bends and fan-out are also possible, as shown in figure 5(b) and (c), respectively. Again, the left-hand cell is fixed and the ground-state configuration calculated. This

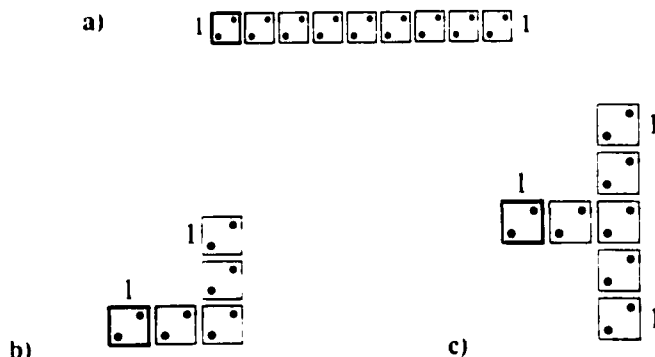


Figure 5. QCA wires: (a) the basic wire; (b) a corner in a wire; (c) fan-out of one signal into two channels. In each case the darker (left-hand) cell has a fixed polarization which constitutes the input. Note that these figures are not simply schematic, but are a plot of the results of a self-consistent many-body calculation of the ground state for the cellular array. The diameter of each circle is proportional to the calculated charge density at each site.

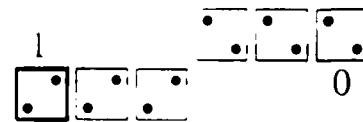


Figure 6. An inverter constructed from a quantum cell automaton

sort of fan-out is appropriate for the edge-driven paradigm discussed above.

4.2. Inverter

By offsetting one chain of cells from another, as shown in figure 6, an inverter can be constructed. If the polarization of the one end is fixed, the polarization of the other end will be opposite.

4.3. AND and OR gates

AND and OR gates can be made from the intersection of two wires. Figure 7 shows an OR gate. The darker boxes are around the input cells. Their polarization is set to correspond to the logical values shown. For the case when the inputs are 0 and 1 (figure 7(c)), the central cell state would normally be indeterminate since a 'tie vote' exists between the input cells. To resolve, this we bias the central cell by increasing the site energy on sites 2 and 4 slightly. This could be accomplished by making the quantum dot diameter slightly smaller on these two sites. It is then slightly more energetically favorable for the cell to be in a 1 state, thus breaking the tie. The AND gate is constructed in exactly the same way except that the central cell is biased toward the 0 state. The AND gate is shown in figure 8. Both these figures reflect the results of self-consistent solutions of the many-electron problem for the entire array shown.

4.4. Memory cell

A single quantum cell can act as a memory storage cell. Once prepared in an eigenstate with $P = +1$, for

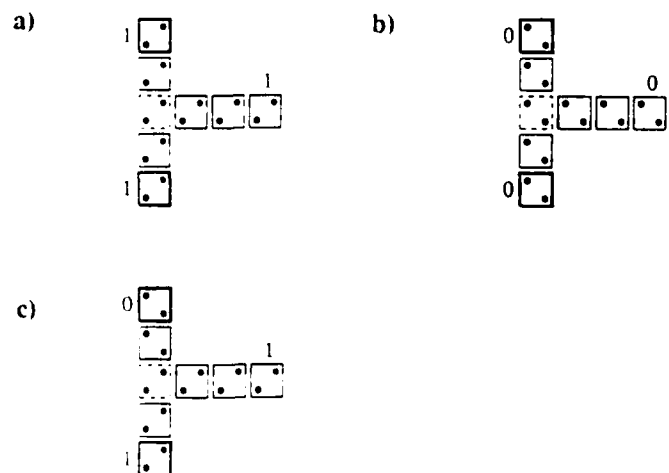


Figure 7. An OR gate. The cells in darker squares are fixed to the input states. The cell in the dashed square is biased slightly toward the '1' state.

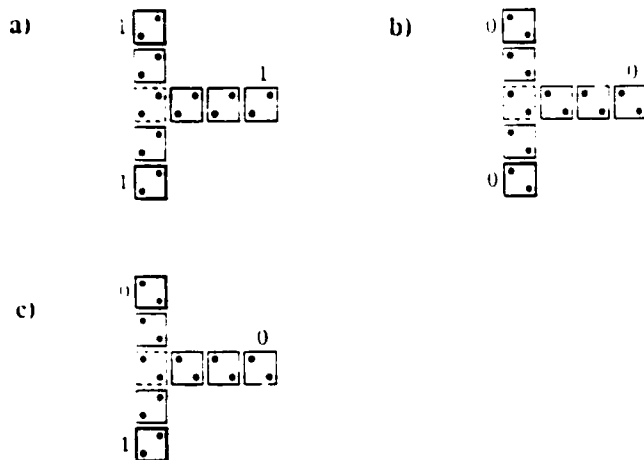


Figure 8. An AND gate. The cells in darker squares are fixed to the input states. The cell in the dashed square is biased slightly toward the '0' state.

example, the cell will in principle remain in that configuration indefinitely. One problem is that slight variations in the potential environment may make it slip into the other eigenstates. To avoid this it may be desirable to use small or medium-size arrays of quantum cells to store each bit. This is shown schematically in figure 9. One advantage of a regular rectangular array of cells is that it may be possible to use the interaction of many cells with the set and sense lines (the exact mechanism for setting and sensing is not critical here). The problem of making non-interfering address lines is certainly non-trivial.

5. Issues for QCA as a technology

Fabrication of QCAs in semiconductors appears to be within reach of current technology. The GaAs/AlGaAs system has proven fruitful as a means of fabricating quantum dot structures by imposing electrostatically a pattern on the two-dimensional electron gas formed at the heterojunction interface. Other materials systems, including molecular systems, are also candidates for

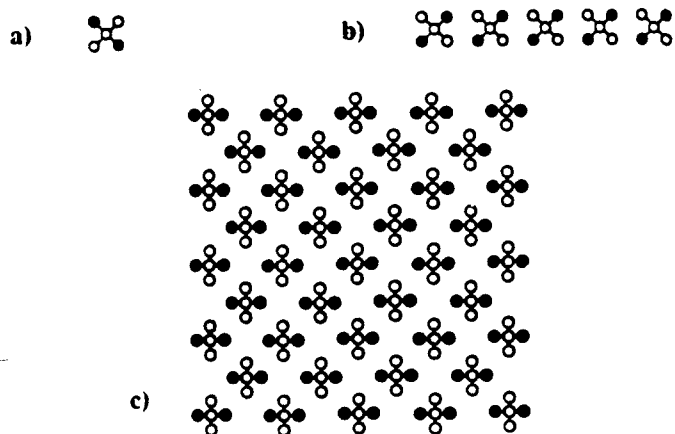


Figure 9. Quantum cellular arrays as memory storage cells. A single bit can be stored in (a) a single cell, (b) a line of cells, or (c) an array of cells. Arrays of cells would make the storage more robust.

realizing a QCA structure. Any implementation must deal with several issues of key importance to the successful operation of the cell we have described.

5.1. Uniformity of cell occupancy

It is important for the operation of the QCA that each cell contain two electrons. The cell-cell response function degrades significantly if one or three electrons are in a cell. Fortunately, the physics of the cell acts to ensure that the occupancy will be very uniform. This is so because the Coulomb interaction causes significant energy-level splitting between the different cell charge states. The Coulomb energy cost to add the third electron is on the order of 10 meV for cells with a 30 nm separation. Experiments by Meurer *et al* [13] have shown that uniformity in the number of electrons per dot can be maintained in arrays of 10^8 dots.

5.2. Dot size control

The size of the fabricated quantum dots must be fairly well controlled. Variations in the size of the dots translate into variations in the confinement energies on each dot. The cell bistability occurs because the Coulomb interaction is determinative in selecting a preferred polarization state. If the magnitude of the variation among the dots in the confinement energies is greater than the Coulomb energies involved, the cell will be pinned at a fixed polarization. Note that dot size variations are critical only within a single cell; variations between different cells are easily tolerated.

5.3. Temperature

The temperature of operation is a major factor. Our QCA quantum cell is expected to work at liquid helium temperatures for dot dimensions which are within the capability of current semiconductor fabrication technology. As technology advances to smaller and smaller dimensions on the few-nanometer scale, the temperature of operation will be allowed to increase. Perhaps our envisioned QCA will find its first room temperature implementation in molecular electronics.

6. Technological benefits

If successful, QCAs would represent a revolutionary, rather than evolutionary, departure from conventional electronics. In this section we review some possible benefits a QCA technology might provide.

Quantum cellular automata solve the interconnection problem. It is widely acknowledged that the main challenge to further improvements in microelectronics is the interconnection and wiring problem. The QCAs we discuss accommodate this challenge in a natural fashion. Interconnect lines are no longer necessary to provide the communication between cells; the Coulomb interaction

provides the coupling mechanism. Edge-driven computation requires neither energy nor information to be transmitted directly to interior cells. Computing with the ground state makes both clocking and refresh signals unnecessary.

Quantum cellular automata make possible ultra-high-density computing elements. The chief technological advantage of the proposed structures is the improved functional density of computing elements. With a 10 nm design rule, the cell dimensions would be about 50 nm × 50 nm, which translates into an extremely high packing density of about 10^{10} cells cm⁻². Since, as shown above, a single cell can function as a logical gate, this represents an extremely high functional density.

Quantum cellular automata are extremely low in power dissipation. High packing density is usually accompanied by high power dissipation. However, in QCA structures, the information is stored in physical systems close to their ground state. The energy input to the array is the energy required to set each input bit—about 1 meV per input bit. This energy is dissipated in the time it takes for the QCA to relax to its new ground-state configuration, probably less than a few picoseconds (phonon scattering times). This represents a power dissipation of roughly 10^{-10} W per input bit, much less than conventional devices.

Quantum cellular automata offer the possibility of ultra-fast computing. As estimated above, the computation occurs in a QCA over the relaxation time for the electrons in the array, probably on the order of picoseconds. It is clear that this relaxation time is a function of the electron-phonon coupling and represents a fundamental speed limit for computation with electrons in a semiconductor.

Quantum cellular automata may facilitate fabrication of ultra-dense memory storage. The QCA cell encodes a bit of information. Writing and reading the bit involves very low power dissipation and is very fast. While problems of cell addressing and cell volatility appear challenging, the possibility of solid-state electronic storage of information at these densities invites further investigation.

7. Summary

We have presented a specific model for using nanoelectronic devices in a cellular automata architecture and proposed a new paradigm for computing in this framework. Each cell consists of a central quantum dot and four neighboring dots occupied by two electrons. The Coulomb repulsion between the two electrons, quantum confinement effects, and the discreteness of the electronic charge, combine to produce strongly polarized (in the

sense defined above) ground states. The response of this polarization to the electrostatic environment is highly non-linear and exhibits the bistable saturation necessary for a two-state CA. The concept of *edge-driven computation* solves the interconnection problem. The concept of *computing with the ground state* in the QCA approach permits ultra-fast operation, eliminates problems of interconnect delays, resistive and capacitive effects, power dissipation, and limited densities associated with conventional architectures.

Acknowledgments

This work was supported in part by the Air Force Office of Scientific Research and the Office of Naval Research. This material is based upon work supported under a National Science Foundation Graduate Fellowship.

References

- [1] Ferry D K, Akers L A and Greeneich E W 1988 *Ultra Large Scale Interconnected Microelectronics* (Englewood Cliffs, NJ: Prentice Hall)
- [2] For a recent overview see Kirk W P and Reed M A (ed) 1992 *Nanostructures and Mesoscopic Systems* (Boston: Academic)
- [3] Bate R T 1977 *Bull. Am. Phys. Soc.* **22** 407
- [4] Randall J N, Reed M A and Frazier G A 1989 *J. Vac. Sci. Technol.* **B7** 1398
- [5] Ferry D K, Barker J R and Jacoboni C (ed) 1991 *Granular Nanoelectronics* (New York: Plenum)
- [6] Landauer R 1989 *Phys. Today* **42** 119
- [7] A proposal for cells with single-electron occupancy has been made, but lacks the requisite bistable character. See Bakshi P, Broido D A and Kempa K 1991 *J. Appl. Phys.* **70** 5150
- [8] Lent C S, Tougaw P D and Porod W 1993 *Appl. Phys. Lett.* **62** 714
- [9] The edge-driven paradigm proposed here is to be distinguished from conventional systolic architectures. In systolic arrays *information* is input only at the edges, but *energy* must be separately fed to each computational element, typically through power lines to each cell. In the paradigm discussed here, both energy and information are supplied only to the edge cells.
- [10] Obermayer K, Mahler G and Haken H 1987 *Phys. Rev. Lett.* **58** 1792
- [11] A one-dimensional treatment is given at length in Morrison M, Estle T and Lane N 1976 *Quantum States of Atoms, Molecules, and Solids* (Englewood Cliffs, NJ: Prentice-Hall) ch 13
- [12] Toffoli T and Margolus N 1987 *Cellular Automata Machines: A New Environment for Modeling* (Cambridge, MA: MIT Press)
- [13] Meurer B, Heitmann D and Ploog K 1992 *Phys. Rev. Lett.* **68** 1371

A novel quantum wire formed by lateral p - n - p junctions between quasi-two-dimensional electron and hole systems at corrugated GaAs/AlGaAs interfaces

Henry K. Harbury and Wolfgang Porod

Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

Stephen M. Goodnick

Department of Electrical and Computer Engineering, Oregon State University, Corvallis, Oregon 97331

(Received 6 August 1992; accepted for publication 19 October 1992)

Numerical modeling of a novel quantum wire structure formed by the confinement of electrons between lateral quasi-two-dimensional (Q2-D) p - n junctions in a corrugated GaAs/AlGaAs heterostructure is reported on. Such a quantum wire may be realized at the tip of a Si-doped AlGaAs overgrown V groove in a Si-GaAs substrate due to the surface orientation dependence of Si doping. The two-dimensional conduction and valence band potential profiles for the electron and hole charge densities are solved within a semiclassical Thomas-Fermi screening model. The quantized electronic wire states at the heterointerface are then obtained by solving the two-dimensional effective mass Schrödinger equation using the calculated potential profile. The parameter space of the one-dimensional electron system is explored to establish which features of the structure are dominant factors in controlling the electronic states. It is demonstrated that the energy level spacing of the quantum wire depends primarily on the lateral confinement width in the n -type region at the tip of the V groove. The ground state energy of the wire is shown to depend on both the lateral confinement width and the vertical heterointerface confinement width. The results of our initial calculations are also reported on to incorporate lateral gates on the surface to obtain direct control of the quantum wire transport properties. The advantages of fabricating quantum wires with this structure compared to conventional methods of electrostatic confinement are discussed.

1. INTRODUCTION

Recent experiments have demonstrated that Si behaves as a substrate orientation dependent amphoteric dopant in molecular beam epitaxial (MBE) grown GaAs and AlGaAs. This effect was first studied on As-terminated {111} and {110} planes;¹ then further studies were reported on Ga-terminated {111}A, As-terminated {111}B, {100}, {211}, and higher index planes.^{2,3} These studies show that, under suitable growth conditions, Si-doped overlayers grown on {100} GaAs substrate surfaces have donor behavior, whereas Si-doped overlayers grown on Ga-terminated {111}A GaAs substrate surfaces have acceptor behavior.

The controllable amphoteric nature of Si doping incited the investigation of fabricating lateral p - n junctions in GaAs. Bulk GaAs lateral p - n junctions with good diode current-voltage characteristics were reported by Miller in 1985.⁴ HEMT-compatible heterostructure lateral p - n junctions, between quasi-two-dimensional electron and hole systems, were reported by Ebner *et al.* in 1990.⁵ In this latter study, Si-doped $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ was epitaxially regrown on a semi-insulating corrugated GaAs substrate. The selectively etched and epitaxially regrown structure was fabricated with exposed Ga-terminated {111}A facets on the normally exposed {100} substrate. Ebner *et al.*⁵ reported electroluminescence results that correspond to the GaAs quantum well band gap of such a system, and Harbury *et al.*⁶ reported calculations that confirm the existence of a lateral p - n junction between two-dimensional electron and

hole gas systems (2DEG and 2DHG, respectively) with Si-doping densities in excess of $5 \times 10^{17} \text{ cm}^{-3}$.

A variety of novel complementary device structures may be based on lateral p - n junctions formed between high mobility 2DEG and 2DHG systems. In particular, exploiting the amphoteric dopant behavior of Si in a selectively etched and epitaxially regrown V groove structure presents a novel quantum wire system realized between lateral p - n junctions.⁷ The depletion regions in such a HEMT-compatible p - n - p type geometry provide lateral confinement of the carriers at the heterointerface of the n -type center region, thereby realizing a quantum wire. The use of p - n junctions for lateral confinement of quantum wires was pioneered by Dean and Pepper⁸ and by Fowler and co-workers.⁹ V -groove geometries have been previously explored for the design of quantum wires.¹⁰⁻¹³ This new design, however, may offer some advantages compared to present unipolar structures,¹⁴ which use metal gate electrostatic confinement or sidewall etching.

In this paper we report on numerical modeling of the lateral p - n - p quantum wire system realized in a corrugated GaAs/AlGaAs heterostructure. In Sec. II we present our model of the system, and in Sec. III we discuss our solution methods for obtaining the two-dimensional electrostatic potential for electrons and holes, and the electronic states of the quantum wire. We also present the details of the numerical method in the Appendix. Our results are presented in Sec. IV, and we conclude in Sec. V.

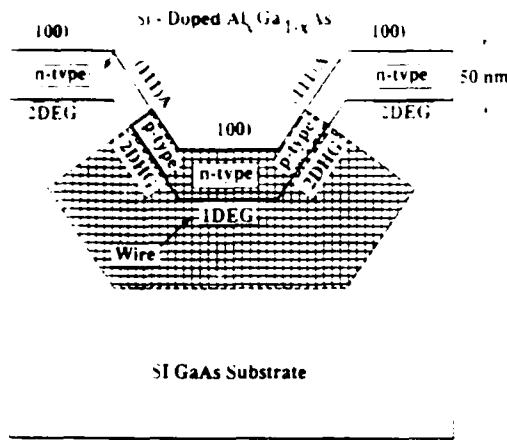


FIG. 1. Schematic diagram of the model V -groove corrugated heterostructure geometry. The n -type or p -type nature of the amphoteric silicon doped overlayer is labeled along with the corresponding crystallographic surface orientation. Possible formation of 2DEG's, 2DHG's, and a 1DEG is also noted. The hatched region represents the computation domain.

II. MODEL SYSTEM

Figure 1 shows a schematic diagram of the model corrugated GaAs/AlGaAs heterojunction system. The chosen overlayer thickness and composition are similar to those of a lateral light emitting structure fabricated by Ebner *et al.*⁵ A 50 nm thick layer of Si-doped $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ is overgrown on an etched V groove of a semi-insulating GaAs substrate. The sidewalls of the V -groove expose the $\{111\}$ A family of planes of the normally $\{100\}$ terminated substrate. As indicated in Fig. 1, the amphoteric Si-doped overlayer exhibits n -type behavior on the $\{100\}$ surfaces and p -type behavior on the $\{111\}$ A surfaces. The two-dimensional electron and hole gas shown in Fig. 1 (labeled 2DEG and 2DHG, respectively) forms along the heterointerface for suitable doping conditions.

The hatched region in Fig. 1 represents the finite element calculation domain used to model the p - n - p quantum wire structures. The range of the ternary composition, the doping density, and the structural dimensions are the parameters that define the system. A two-dimensional equilibrium potential profile of the model geometry is sought for different material parameters. The band bending due to the electrostatic potential, $\phi(r)$, is obtained from the solution of Poisson's equation within a semiclassical Thomas-Fermi screening model,

$$\nabla^2 \phi(r) = \frac{e}{\epsilon \epsilon_0} (n(r) - p(r) + N_A^- - N_D^+). \quad (1)$$

Complete ionization of the impurities is assumed at room temperature, such that the background doping of the semi-insulating GaAs, assumed to be lightly n type, contributes to N_D^+ in the substrate, and the amphoteric Si doping contributes to both N_A^- and N_D^+ in the overlayer region. The electron density, $n(r)$, is given by the Fermi-Dirac integral of order $\frac{1}{2}$,¹⁵

$$n(r) = 4\pi \left(\frac{2k_B T m^*}{h^2} \right)^{3/2} \int_{E_c(r)}^{\infty} \frac{([E - E_c(r)]/k_B T)^{1/2} dE/k_B T}{1 + \exp((E - E_F)/k_B T)} = N_c F_{1/2}(\eta), \quad (2)$$

where $N_c = 2(2\pi k_B T m^*/h^2)^{3/2}$ is the effective conduction band density of states, $\eta = (E_F - E_c(r))/k_B T$ measures the separation between the Fermi level, E_F , and the position-dependent conduction band edge, $E_c(r)$, and $F_{1/2}(\eta)$ is the Fermi-Dirac integral of order $1/2$. A similar term is obtained for the hole density, $p(r) = N_v F_{1/2}(\eta)$, where η is now the separation between the valence band edge and the Fermi level, and N_v is the effective valence band density of states.

We enforce bulk charge neutrality deep inside the semi-insulating GaAs substrate, assumed to be slightly n type with a background doping density of $1.0 \times 10^{15} \text{ cm}^{-3}$ which is completely ionized. A density of surface states along the $\{100\}$ and $\{111\}$ exposed facets is also assumed, such that the electrostatic potential is pinned to the near midgap value of 0.8 eV separation between the conduction band edge and the Fermi level. The problem domain extends far enough from the p - n junction regions, such that the electrostatic potential becomes invariant parallel to the heterointerface.

The confined states of the p - n - p quantum wire structure are also of interest. The bound state wave functions are sought by solving the two-dimensional Schrödinger equation for a computed potential profile,

$$-\frac{\hbar^2}{2m_c^*} \nabla^2 \psi(r) + [V(r) - e\phi(r)] \psi(r) = E \psi(r). \quad (3)$$

Here, $V(r)$ includes external potentials and band offsets and $\phi(r)$ is the electrostatic potential obtained from the solution of Poisson's equation for the V -groove geometry.

The decay of the wave function far from the potential "pocket" that forms the quantum wire provides the necessary boundary conditions to formulate Eq. (3) as an eigenvalue/vector problem. The wave functions of interest, however, are the lowest energy quantum wire states. The problem, therefore, can be reduced to a subspace of the full eigensystem.

III. NUMERICAL METHODS

The solution of the two-dimensional Poisson and Schrödinger equations is a challenging numerical problem that requires domains of highly nonuniform mesh design, bandwidth optimization, and sparse matrix methods. The general considerations of the solution method are discussed in Sec. III A, whereas the details of the numerical method are deferred to the Appendix. The interesting and often ignored subject of nonuniform mesh generation is discussed in Sec. III B.

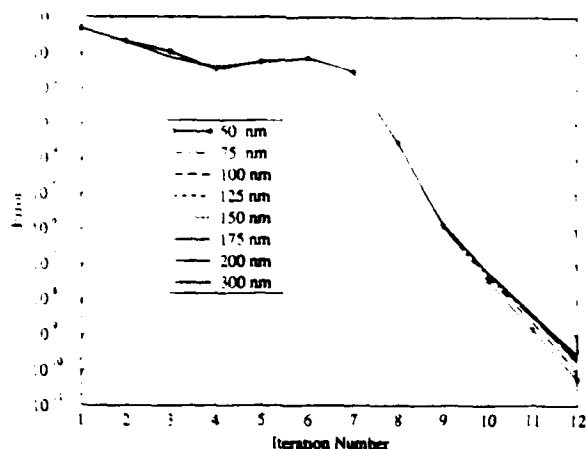


FIG. 2. Convergence of the Poisson solution Newton-Raphson iteration scheme for several different p - n - p geometries.

A. The solution of Poisson and Schrödinger equations

The two-dimensional Poisson and Schrödinger equations are both solved with the finite element method. The Poisson equation is solved for the discretization domain schematically shown by the hatched region in Fig. 1, whereas the Schrödinger equation is only solved for the small subregion near the V -groove tip, where the electrostatic potential "pocket" defines the quantum wire.

For the Poisson solution, the bulk charge neutrality boundary conditions deep inside the GaAs substrate are implemented by fixing the electrostatic potential at a few points on the boundary to the value consistent with the background doping, and by applying zero-valued Neumann conditions on the remainder of the bulk boundary. This condition forces the boundary normal electric field to zero, and forces the potential to the proper equilibrium bulk value. Surface pinning on the exposed facets is implemented by Dirichlet conditions, which force the conduction band potential to the approximate midgap value of 0.8 eV above the Fermi level. Zero-valued Neumann boundary conditions are applied along the side boundaries to model the asymptotic invariance parallel to the heterointerface. The discontinuity in the electrostatic potential at the heterointerface is treated as a linear constraint on the system of equations and is implemented with the penalty element method. The nonlinear Poisson equation is linearized by the Newton-Raphson iterative method. The details of the finite element formulation and linearization are deferred to Appendix A.

The resulting system of linear equations is solved by standard L/U decomposition, backward substitution, and forward elimination. Bandwidth optimization of the discretized equation numbers in conjunction with the skyline storage technique provide an efficient solution of the linear system. The iterative error of the Newton-Raphson solution is plotted in Fig. 2 for typical runs of the lateral p - n - p quantum wire domain for various widths of the n -type region at the V -groove tip. The finite element domain consists of 16 168 elements and 8427 bandwidth-optimized nodes. The error drops by approximately 9 decades within

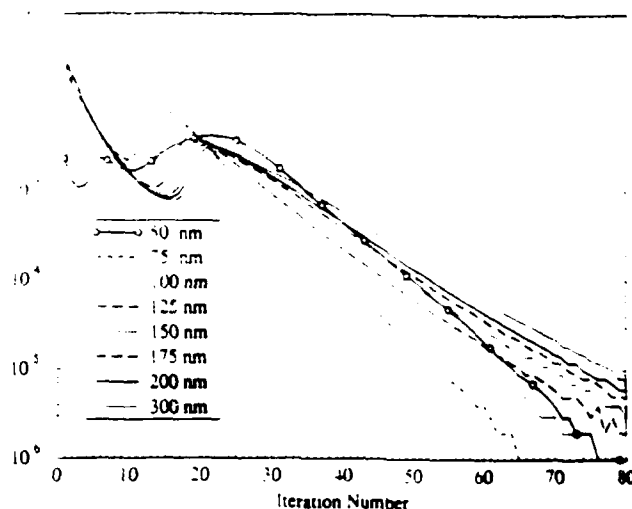


FIG. 3. Convergence of the Schrödinger solution subspace iteration scheme for several different p - n - p geometries.

12 iterations with a typical cumulative solution time of 430 sec on an IBM RISC System/6000.

For the Schrödinger equation, the asymptotic decay of the wave function can be modeled by either zero-valued Neumann boundary conditions or zero-valued Dirichlet boundary conditions. Continuity of the wave function across the heterointerface is implemented with the penalty element method and provides compatibility with the Poisson solver. The details of the finite element formulation of the Schrödinger equation are deferred to Appendix B.

Because only the lowest few eigenstates are sought, the discretized Schrödinger equation is solved by a subspace iteration method. Efficient skyline storage and sparse matrix methods are used to obtain the lowest ten bound states. The iterative error of the subspace algorithm is plotted in Fig. 3 for the ground state of the lateral p - n - p quantum wire for the same selection of n -region V -groove tip widths, as were chosen for Fig. 2. After initial oscillation, the error exponentially decreases at approximately one decade every 15 iterations. For the finite element Schrödinger domain consisting of 3482 elements and 1876 bandwidth-optimized nodes, a solution of a 20 eigenstate subspace over 120 iterations typically consumes 240 cpu seconds on an IBM RISC System/6000.

B. Nonuniform mesh generation

Two-dimensional numerical modeling of realistic structures, such as those considered in this work, requires the development of sophisticated mesh domains. Bulk charge neutrality boundary conditions imply that the domain, Ω , must extend microns into the bulk region. At the heterointerface and in the p - n junction depletion regions, however, the potential profile may undergo large variations in nanometer distances. The disparity between characteristic length scales and the need to keep the problem size computationally tractable constrains the discretization scheme to highly nonuniform mesh designs. A further constraint on mesh design is the need to preserve the charac-

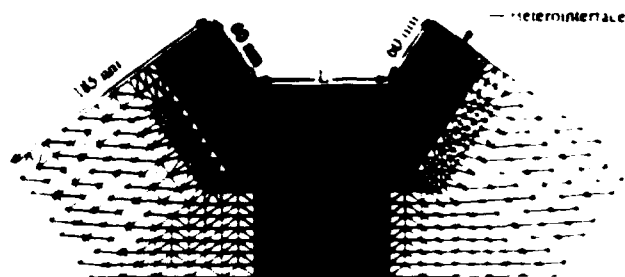


FIG. 4. Nonuniform, triangular element mesh domain used to solve the Schrodinger equation for the quantum wire state. The Schrodinger mesh is a small subdomain of the Poisson mesh cut from the region of the potential "pocket" formed at the bottom of the V groove. This mesh has 3482 triangular elements and 1876 nodes.

teristic geometry of the structure. The 54.7° angle between the $\{100\}$ and $\{111\}$ crystallographic planes, for example, must be maintained throughout the domain.

With a rectangular master element design, it would be difficult to maintain arbitrary angles throughout the domain, and the resulting system of equations would likely be intractably large. With the triangular master element, however, it is easy to maintain arbitrary geometric angles in nonuniform mesh designs. Furthermore, standard bandwidth optimization algorithms exist for triangular element domains to efficiently number the nodes for sparse matrix methods.

At present there are few mesh layout tools available to the scientific community suitable for device modeling applications. Xmgredit,¹⁶ however, is a powerful mesh generation tool that was originally designed for oceanographic land-margin studies, which is general enough for device modeling work. Xmgredit is a window based interactive mesh editing tool that allows arbitrary triangulation of a domain and provides all the necessary functions for grid refinement.

Figure 4 shows the mesh generated to solve the Schrodinger equation for the lateral p - n - p quantum wire structure. This mesh is a small subdomain of the larger mesh used to solve the Poisson equation, since the bound state wave functions are nonzero only near the potential "pocket" at the V -groove tip. We present only the smaller Schrodinger mesh in Fig. 4 because its ratio of largest to smallest element is small enough that all the elements can be clearly seen.

As mentioned above, the Poisson mesh consists of 16 168 elements and 8427 nodes, and the Schrodinger subdomain consists of 3482 elements and 1876 nodes. For the larger Poisson mesh, the maximum element area is $3.7 \times 10^5 \text{ \AA}^2$ and the minimum element area is 70 \AA^2 , while maintaining the proper crystallographic orientations throughout the domain. The sidewalls of the V groove extend approximately 800 nm for the Poisson mesh and approximately 60 nm in the Schrodinger mesh, as shown in Fig. 4.

In designing such a large nonuniform mesh, it is easiest to generate a separate mesh for each region of the heterostructure with different material properties and then con-

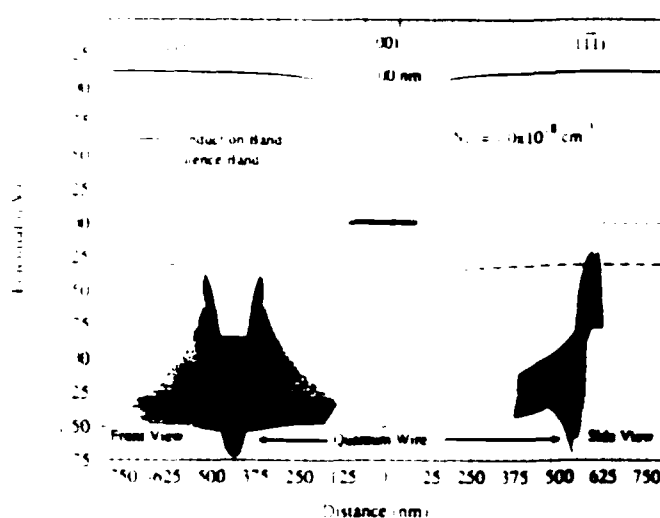


FIG. 5. Band diagram for the lateral p - n - p quantum wire structure. Shown are the conduction and valence bands along the heterointerfaces for a 100 nm long center n -type section and a Si-doping density of $1.0 \times 10^{18} \text{ cm}^{-3}$. The insets show front and side views of the two-dimensional conduction band profile. The potential "pocket" that holds the quantum wire is also indicated.

nect the subdomains together. The special penalty elements can then be added along the heterointerfaces to force the potential discontinuity. We have found that a computationally efficient triangular element mesh can be generated for any two-dimensional device geometry with a minimum of effort.

IV. LATERAL V-GROOVE QUANTUM WIRE STRUCTURE

The results presented in this section concentrate on the p - n - p V -groove structure indicated by the hatched model domain labeled "wire" in the schematic diagram shown in Fig. 1. Interesting behavior is expected when back-to-back lateral quasi-2-D p - n junctions are formed with a common n -type region. In particular, if the n -type region is made narrow, one would expect the 2DEG formed at the heterointerface to be laterally "squeezed" into a 1DEG by the depletion regions of the p - n junctions on both sides. A conduction band "pocket" is expected to form at the tip of the V groove, which may support two-dimensionally confined quantum wire states.

Figure 5 is a plot of the band diagram for a lateral p - n - p junction with an overlayer Si-dopant concentration of $1.0 \times 10^{18} \text{ cm}^{-3}$ and with a narrow 100 nm n -type region at the V -groove tip. The electrostatic potential is plotted parallel to the heterointerface on the GaAs substrate side, along the $\{111\}$ - $\{100\}$ - $\{111\}$ surfaces. The formation of an electron gas is expected in the region where the conduction band approaches and traverses the semiclassical Fermi energy, indicated by the dashed line, which is chosen as the zero of energy. The quantum wire potential "pocket" is more clearly visible in the insets to Fig. 5, where portions of the front and side views of the two-dimensional conduction band profile are shown. Only a small submesh of the full two-dimensional computation domain is shown in the

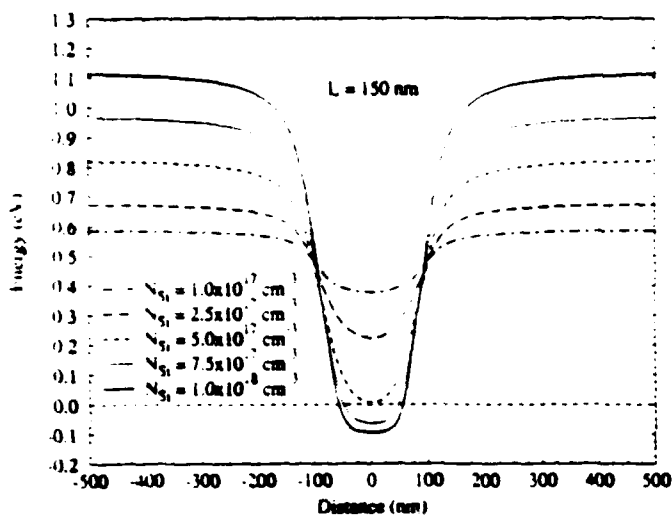


FIG. 6. Conduction band profiles for the p - n - p quantum wire structure. Shown are the conduction bands along the (111) and (100) heterointerfaces with a fixed 150 nm width n -type section for five different overlayer dopant concentrations.

inset figures for better delineation of the potential well, although the calculations are performed on the full domain to ensure that bulk charge neutrality conditions are satisfied deep in the GaAs substrate. The lateral p - n - p junctions that form at the heterointerface are clearly visible in the left inset of Fig. 5 (front view), whereas the heterointerface and "pocket" are more clearly seen in the right inset (side view).

Shown in Fig. 6 is a comparison of the lateral conduction band profile for several different overlayer Si-dopant concentrations with constant 150 nm width n -type regions. Again, the semiclassical Fermi energy is chosen as the zero of energy and is indicated by the dashed line. As is expected, the p - n - p depletion widths are reduced and the quantum well depth in the n region increases with higher overlayer dopant concentration. For the case of a 150 nm wide V -groove tip and overlayer dopant concentration of $1.0 \times 10^{18} \text{ cm}^{-3}$, flatband conditions are approached at the center of the n region, as shown by the solid curve in Fig. 6. This flatband condition occurs when the two lateral depletion regions do not overlap and therefore there is no lateral field component in the center (n -type) region of the V -groove tip.

The effect of the width of the n -type V -groove tip on the lateral conduction band profile is shown in Fig. 7 for a constant Si overlayer dopant concentration of $1.0 \times 10^{18} \text{ cm}^{-3}$. The semiclassical Fermi energy is again indicated by the dashed line at 0.0 eV. Larger V -groove tip widths correspond to an increase in depth and width of the lateral quantum well. Flatband conditions are demonstrated in Fig. 7 for n -region widths greater than 150 nm. As in Fig. 6, lateral flatband conditions occur at the bottom of the quantum well "pocket" when the two lateral p - n depletion regions do not overlap. Once flatband conditions are established in the bottom of the well, the maximum depth of the "pocket" becomes independent of the V -groove tip width. However, the electronic states in the well are expected to be a strong function of the lateral well width. In the limit

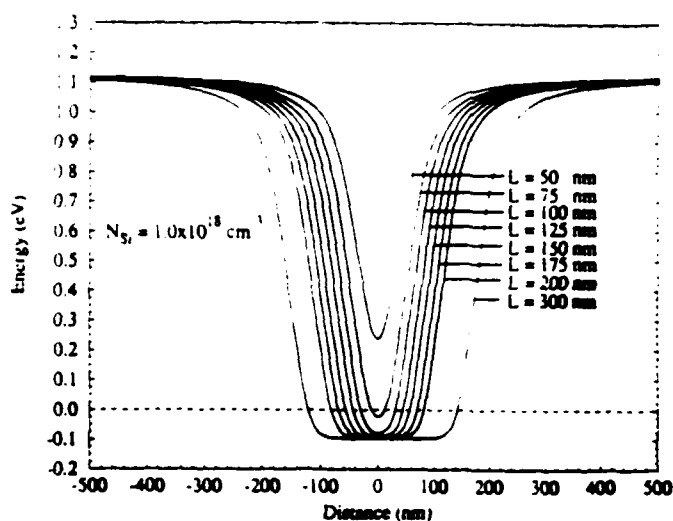


FIG. 7. Conduction band profiles for the p - n - p quantum wire structure. Shown are the conduction bands along the (111) and (100) heterointerfaces with a fixed $1.0 \times 10^{18} \text{ cm}^{-3}$ doped $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ overlayer for eight different widths of the exposed (100) surface.

of a very wide n region, the behavior of the system will approach that of isolated p - n junctions and the quasi-1DEG becomes a quasi-2DEG in the n region. The V -groove tip width, therefore, controls the transition of the n region from a 1DEG quantum wire system to a 2DEG lateral p - n - p system.

Figures 6 and 7 demonstrate how control of the lateral quantum well electrostatic potential can be obtained through adjusting both the overlayer Si-dopant concentration and the lateral V -groove dimensions. It is also of interest to study the dependence of the vertical heterointerface confinement for the same parameters. Plotted in Fig. 8 are the conduction band profiles in a direction perpendicular to the heterointerface, at the center of a 150 nm wide n region, for the same dopant conditions used in Fig. 6. As before, the zero of energy is chosen to be the semiclassical Fermi energy and is indicated by the dashed line. Increasing negative values of distance correspond to increasing depth into the GaAs substrate, whereas increasing positive

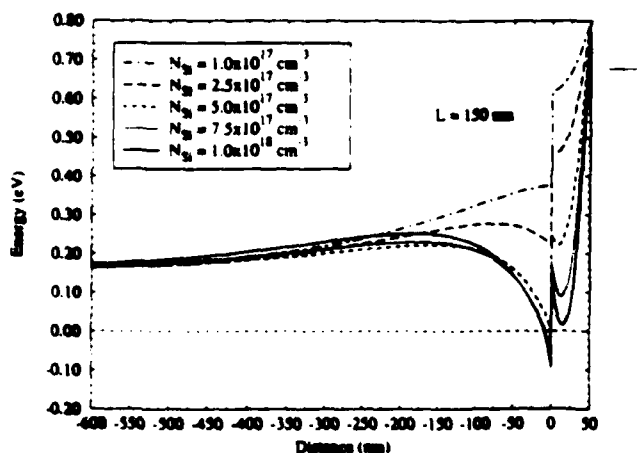


FIG. 8. Potential variation perpendicular to the heterointerface (at the center of a fixed 150 nm width n -type region) for five different overlayer dopant concentrations.

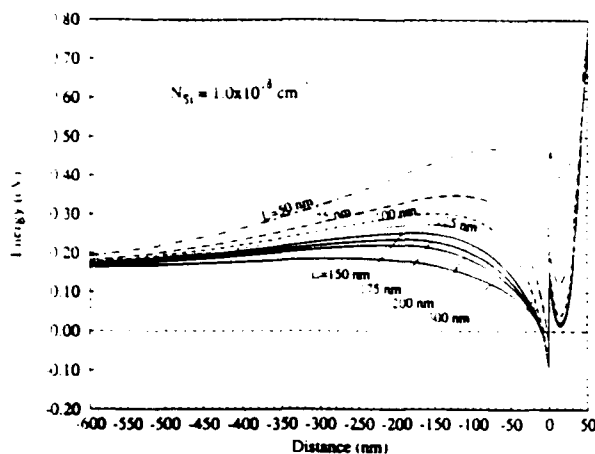


FIG. 9. Potential variation *perpendicular* to the heterointerface (at the center of the *n*-type region) with a fixed $1.0 \times 10^{18} \text{ cm}^{-3}$ Si-doped $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ overlayer for eight different widths of the *n*-type region.

values of distance correspond to locations in the $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ overlayer approaching the surface. For clarity, the results are only plotted up to a depth of 600 nm into the GaAs substrate, although the calculation domain extends for several microns to ensure that bulk charge neutrality conditions are satisfied. Figure 8 shows that heterointerface confinement is only achieved for overlayer dopant concentrations above a critical value of approximately $5.0 \times 10^{17} \text{ cm}^{-3}$. The conduction band potential on the GaAs side at the heterointerface has a strong dependence on the overlayer doping for concentrations below the critical value, marking the onset of the electron gas formation, and a much weaker dependence for concentrations above this critical value. The conduction band potential in Fig. 8 first *increases* from its bulk value for points approaching the heterointerface, and then it sharply decreases to form the heterointerface confinement "notch." This "hump" in the GaAs conduction band is more pronounced for higher overlayer dopant concentrations.

In a similar fashion to Fig. 8, Fig. 9 shows the corresponding conduction band profiles for the same selection of center *n*-region widths used in Fig. 7 with a constant overlayer dopant concentration of $1.0 \times 10^{18} \text{ cm}^{-3}$. The most noticeable feature is the pronounced "hump" in the GaAs conduction band at narrower *n*-region widths. For the narrowest, $L = 50 \text{ nm}$ case denoted by the dash-dotted line in Fig. 9, a sharp heterointerface confinement notch is formed with a wide "hump" barrier that extends several hundred nanometers into the GaAs. In this case, the barrier in the GaAs should prevent the leaking of any one-dimensional quantum wire states into the bulk substrate. As the *n*-region width is increased, the GaAs barrier decreases until, in the limit, the conduction band profile approaches that of the far *n* side of an isolated *p*-*n* junction.

It is of interest to verify that the quantum wire structure supports two-dimensionally confined electronic states. To this end, the calculated conduction band profile in a region enclosing the potential "pocket" is used as the electrostatic potential in the Schrödinger Hamiltonian, as discussed in Sec. III A. Presented in Fig. 10 are the lowest

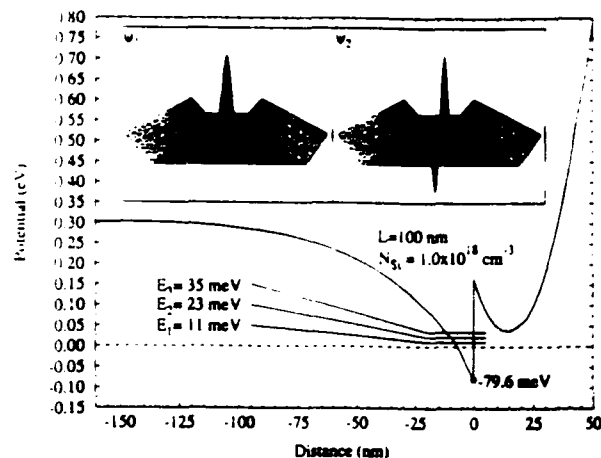


FIG. 10. Potential variation *perpendicular* to the heterointerface (at the center of a 100 nm wide *n*-type region) for the two-dimensional conduction band profile shown in Fig. 5. The insets show the wave functions for the lowest two quantum wire states, that are confined at the GaAs side of the heterointerface (the darker top portion of the mesh corresponds to the AlGaAs overlayer).

energy quantum states sought for the system with a 100 nm *n*-region width and $1.0 \times 10^{18} \text{ cm}^{-3}$ overlayer dopant concentration. The conduction band potential is plotted for a slice through the center of the *n*-type region in a direction *perpendicular* to the {100} surface. The lowest eigenenergy, labeled by E_1 in Fig. 10, is at 11 meV referenced to the semiclassical Fermi energy, which is denoted by the dashed line at 0.0 eV. The subsequent excited states each have a separation of about 12 meV associated with the contribution of the lateral confinement. The left and right insets in Fig. 10 show the full two-dimensional results for the ground state and first excited state, respectively. The $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ overlayer appears as the region of higher mesh density near the top of the inset figures and delineates the heterointerface. The wave functions are concentrated in the deep "pocket" on the higher mobility semi-insulating GaAs substrate side of the heterointerface.

It is clear from the insets of Fig. 10 that the first excited state corresponds to the first excited mode of the *lateral* confinement potential induced by the *p*-*n* junctions. The quantized level spacing for the lower eigenstates is therefore controlled by the nature of the *lateral* confinement. The nearly equal level spacing of the lowest states is due to the parabolic-like shape of the lateral potential profile of the 100 nm width *n*-region structure, as can be seen in Fig. 7. The higher-energy states, however, also include excited modes of the confinement *perpendicular* to the heterointerface. This is particularly true for narrow *n*-region *V*-groove structures in which the magnitude of both confinement dimensions are of similar order. The large separation of the lowest state above the conduction band minima at the heterointerface, labeled by the dot in Fig. 10, is attributed to both the lateral confinement energy and the confinement energy *perpendicular* to the heterointerface.

The details of the amphoteric nature of the silicon doping near the intersection of the (100) and (111) surfaces are not known, and are likely to be more complicated than our model. This and possible compensation effects might

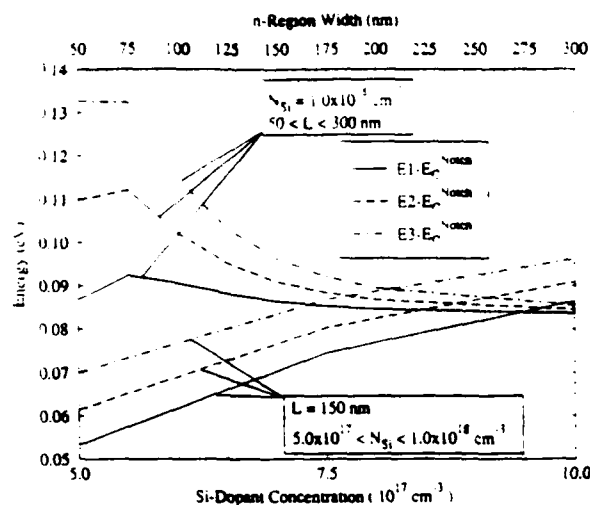


FIG. 11. Variation of the lowest three quantum wire states with overlayer dopant concentration and n -region width. The energies are referenced to the conduction band minimum at the heterointerface (i.e., the bottom of the "pocket"). The negative sloping curves correspond to independently changing the width of the n -type region for a fixed $1.0 \times 10^{18} \text{ cm}^{-3}$ overlayer doping. The positive sloping curves correspond to independently changing the overlayer doping for a fixed 150 nm wide n -type region.

tend to reduce the abruptness of the lateral p - n junction, which also would tend to reduce the energy level spacings.

The dependence of the eigenstates on both the overlayer doping and the n -region width is presented in Figs. 11 and 12. Simultaneously shown in Fig. 11 are the dependence of the lowest three bound state energies on the lateral n -region width and the overlayer Si-doping concentration. The bottom axis corresponds to independently

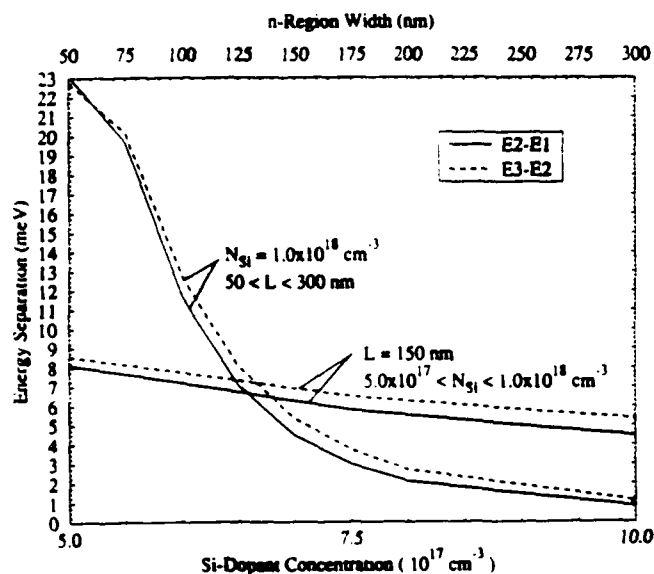


FIG. 12. Variation of the quantum wire state separations with overlayer dopant concentration and n -region width. The sharply sloped curves correspond to independently changing the width of the n -type region for a fixed $1.0 \times 10^{18} \text{ cm}^{-3}$ overlayer doping. The slightly sloped curves correspond to independently changing the overlayer doping for a fixed 150 nm wide n -type region. The solid and dashed lines are the separation between the second and first states and between the third and second states, respectively.

changing the overlayer Si-dopant concentration for a fixed 150 nm width V -groove tip. The top axis corresponds to independently changing the lateral width of the V -groove tip for a fixed overlayer dopant concentration of $1.0 \times 10^{18} \text{ cm}^{-3}$. The energies on the ordinate are plotted relative to the minimum of the conduction band profile at the heterointerface, i.e., relative to the bottom of the potential well, denoted by E_c^{Notch} . The ground state energy is lowered by independently widening the n region and saturates at the limiting value given by the heterointerface confinement. In approaching the wide n -region limit, the one-dimensional electron gas evolves into a two-dimensional electron gas, which is manifest by the pronounced decrease in the level spacing in Fig. 11. The change in behavior for narrow widths, i.e., below 75 nm, seen in Fig. 11, occurs when the magnitude of the lateral junction confinement is comparable to that of the heterointerface confinement. Independently increasing the overlayer dopant concentration has several effects on the potential well; the confining well in the direction *perpendicular* to the heterointerface becomes "deeper" and "thinner," whereas the *lateral* junction confinement becomes wider and more abrupt, with smaller side depletion regions. Figure 11 shows that the net effect of increased doping is to increase the ground state energy, relative to the bottom of the quantum well, with only a slight decrease in the energy level spacings. Simultaneously plotted in Fig. 12 are the energy level separations for the same parameters given in Fig. 11. The bottom and top axis represent the effect of independently changing the overlayer doping and independently changing the n -type region width, respectively. The nearly equal separation of the second mode from the first and the third mode from the second is due to the parabolic shape of the lateral junction confinement. When the lateral n -region width is increased, the pronounced drop in the level spacing is observed, whereas only a slight drop is observed for increased overlayer doping. The n -region width is therefore the dominant controlling parameter of the quantum wire energy level spacing.

Control of the optical and transport properties of the active wire region is of primary interest for eventual device applications of this corrugated GaAs/AlGaAs heterojunction system. The previous results indicate that independent control of the quantum confinement dimensions may be obtained by the fabrication of gates on the corrugated surface. In this section we also present the method and results for our initial attempt to model the electrostatic potential profile with the addition of Schottky gates.

A simple gate geometry is chosen such that each wall of the V groove has a Schottky gate. The gates on the p -type {111} sidewalls stop 100 nm from the n -type {100} V -groove tip. The center gate, along the {100}, extends over the full n -region width. In these calculations, the same bias will be applied to the sidewall gates and the center gate will be referenced to ground.

The Thomas-Fermi screening model for the electron and hole concentrations, described in Sec. II, tacitly assumes an equilibrium carrier distribution. This model is valid for small perturbations from equilibrium, such that

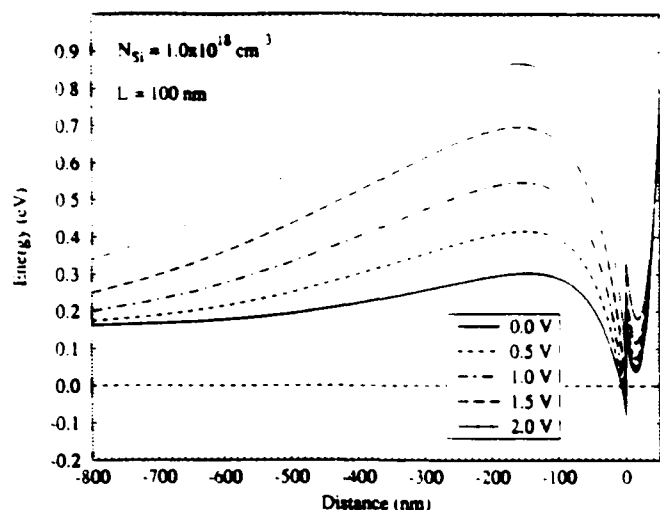


FIG. 13. Potential variation *perpendicular* to the heterointerface (at the center of a 100 nm wide *n*-type region, and $1.0 \times 10^{18} \text{ cm}^{-3}$ Si overlayer doping) for several different bias potentials applied to Schottky gates on the *V*-groove sidewalls.

no appreciable hole concentration exists at the surface. However, under the bias conditions necessary to obtain control of the quantum wire, this charge model must be modified to extend its range of validity. To this end we assume two constant quasi-Fermi levels: one for electrons referenced to the Fermi level in the bulk GaAs substrate, and one for holes referenced to the Fermi level of the sidewall metal gates.¹⁷ The electron and hole concentrations are therefore calculated most accurately in the regions, where they are the majority carrier, and least accurately in the minority carrier regions, where their contribution to the total charge density is least significant.

The boundary conditions on the corrugated AlGaAs surface are modified to include the gate bias. The boundary beneath the gates is modeled by Dirichlet conditions that specify the potential as the sum of the equilibrium surface pinned value plus the bias potential. Since the center gate is referenced to ground, the boundary along the *n*-type region remains pinned at the approximate midgap potential of 0.8 eV above the bulk quasi-Fermi level. The 100 nm exposed surfaces between the gates is modeled by a linear drop in the potential along the surface from the sidewall gates to the center electrode. This somewhat *ad hoc* assumption is justified because the overlayer is depleted of mobile carriers, as seen in the previous results.

The vertical conduction band profile through the center of the *n* region is plotted in Fig. 13 for five different side-gate biases. The overlayer Si dopant concentration is $1.0 \times 10^{18} \text{ cm}^{-3}$ and the *n*-region width is 100 nm. As before, increasing negative values of distance correspond to positions deeper in the bulk GaAs substrate referenced to the heterointerface. The quasi-Fermi level for electrons is specified by the bulk GaAs Fermi level and is denoted by the dashed line at 0.0 eV. Figure 13 demonstrates that, with this gate geometry, increasing bias reduces the confinement width of the notch and only slightly depletes the electron gas. The increase of the "hump" in the GaAs is a

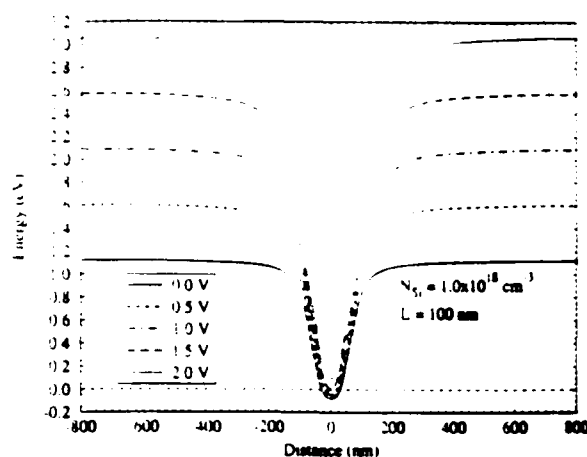


FIG. 14. Lateral conduction band profiles for the same structure and bias conditions, as shown in Fig. 13. Shown are the conduction bands at the heterointerface parallel to the surface along the (100) and (111) planes.

feature that should further confine the electronic states to the heterointerface.

The lateral *p-n-p* conduction band profile is shown in Fig. 14 for the same structure and sidegate bias conditions. Positions parallel to the heterointerface are referenced to the center of the *n* region, and the quasi-Fermi level for electrons is denoted by the dashed line at 0.0 eV. Increasing sidegate bias narrows the lateral confinement dimension, which was previously shown to control the quantum wire level spacing.

These initial results show that control of the electronic properties of the structure may be obtained by the addition of lithographically defined gates on the surface. It is intuitively clear that better control of the lateral confinement width can be achieved with only the sidegates and no center electrode. It is difficult, however, to formulate well-posed boundary conditions along the exposed surface without a center reference electrode. The nature of the Poisson equation requires Dirichlet, Neumann, or mixed boundary conditions along the exposed surface, none of which are simple without completely *ad hoc* assumptions. It is also of interest to form Ohmic contacts to the two-dimensional carriers in the structure, which should provide control of the transport properties through reverse-biased *p-n* junctions. Ohmic contacts also provide the interesting possibility of minority carrier injection into the quantum wire, which may give rise to new device applications, since this is an inherently bipolar structure. The model of an Ohmic contact, however, must include direct control of the two-dimensional carriers, which implies additional constraints along the heterointerface. We are working on a better model for an exposed surface, containing occupied surface trap states under zero current bias conditions, and the development of well-posed electrostatic boundary conditions.

V. SUMMARY

We have reported on our method for modeling a novel *p-n-p* quantum wire structure formed in a corrugated GaAs/AlGaAs heterostructure. We solve the two-dimensional Poisson equation, within a semiclassical

Thomas-Fermi screening model, for the conduction and valence band profiles. The quantum confined electronic states are subsequently obtained by solving the two-dimensional Schrödinger equation using the previously computed electrostatic potential.

We have demonstrated that it is feasible to realize a bipolar quantum wire structure on a HEMT-compatible corrugated GaAs/AlGaAs interface by exploiting the surface orientation dependent amphoteric nature of Si doping. These results indicate that the quantum wire level separations are controlled by the lateral confinement width, and the parabolic-like shape of the lateral confinement potential gives rise to nearly equal level spacings. It was also shown that the ground-state energy of the wire states can be controlled by the vertical confinement width. Initial results demonstrate that the electronic properties of the structure may be controlled by the addition of gates on the corrugated surface. This structure has several possible advantages compared to sidewall etched or Schottky contact electrostatically confined systems.¹⁴ The definition of the n -region width is controlled by the anisotropic etch time through an optically defined photoresist mask; ion-beam or electron-beam lithography is not required to fabricate the structure. This inherently bipolar heterostructure might be used to inject minority carriers into the quantum wire from the p regions, possibly leading to new device applications. If the p -type and n -type regions are separately contacted, by further nanolithography, the lateral confinement width might be controlled by reverse biasing the p - n junctions and the electron Fermi energy might be controlled via a gate contact on the AlGaAs.

ACKNOWLEDGMENTS

The authors would like to acknowledge fruitful discussions with J. Ebner, D. J. Kirkner, C. S. Lent, J. L. Merz, and T. K. Plant regarding this work. The authors would also like to acknowledge the help of J. Westerink for obtaining mesh generation and bandwidth optimization tools and M. Mueller for creating Dore' graphics programs to display the results. This work has been partially supported by the Office of Naval Research and the Air Force Office of Scientific Research.

APPENDIX A: FORMULATION OF THE POISSON PROBLEM

1. Problem statement

The model structure is partitioned into regions, $\Omega_1, \Omega_2, \dots, \Omega_N$, as shown in Fig. 15, defined by different material parameters, such as Al mole fraction and Si-doping concentration. We want to solve the two-dimensional Poisson equation on the whole domain $\Omega \equiv \Omega_1 \cup \Omega_2 \cup \dots \cup \Omega_N$. The total boundary of domain Ω will be denoted by Γ and is partitioned into three terms: Γ_S for that part of Γ that is on the surface of the structure, i.e., the boundary on which the Fermi level is surface pinned, $\hat{\Gamma}$ for that part of Γ on which bulk charge neutrality conditions apply, and last Γ_B for the remainder of Γ on which the normal electric field is forced to zero. Bulk charge neutrality can also be specified

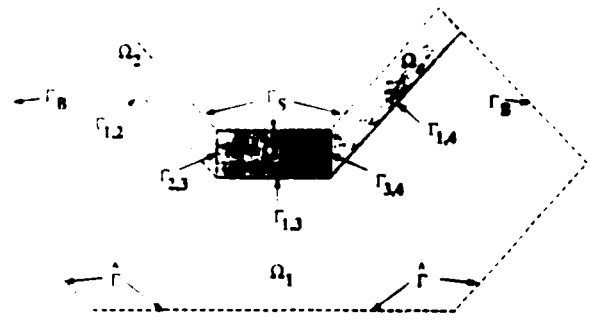


FIG. 15. Schematic diagram of the calculation domain. Domains of different material composition are indicated by different shades of gray and labeled Ω_1 – Ω_4 . The bulk substrate boundary is denoted by $\hat{\Gamma}$, the surface pinned boundary is denoted by Γ_S , and the side boundaries are denoted by Γ_B . The boundaries between adjacent materials Ω_i and Ω_j is denoted by Γ_{ij} . In the quantum wire structure, the AlGaAs/GaAs heterointerface occurs along Γ_{12} , Γ_{13} , and Γ_{14} .

as zero-valued Neumann boundary conditions. We have found that specifying essential boundary conditions at a few points and zero-valued Neumann conditions on the remainder of the bulk boundary works best, since it provides a quick visual check of the solution for proper convergence. The boundaries of each subregion are denoted by Γ_m . The portion of Γ_m in common with another subdomain, Γ_n , is denoted by $\Gamma_{m,n}$. All the heterointerfaces in Ω will therefore occur along the $\Gamma_{m,n}$ boundaries. The potential on the Γ_m side of a heterointerface formed between regions Ω_m and Ω_n is labeled $u_{\Gamma_{m,n}}$, and likewise labeled $u_{\Gamma_{n,m}}$ on the opposite side. We will also define $u = (E_c(x,y) - E_F)/k_B T$ as the conduction band edge energy relative to the Fermi energy in units of $eV/k_B T$. We now specify the zero of potential energy to correspond to the Fermi energy, $E_F = 0$. The problem to be solved can now be written as follows:

Find:

$$u = u_1 \cup u_2 \cup \dots \cup u_N \in H^2(\Omega), \quad (A1)$$

such that

$$-\nabla^2 u_m(x,y) = \frac{\rho(x,y)}{\epsilon_m \epsilon_0}, \quad (x,y) \in \Omega_m \quad (A2)$$

and

$$u_m = \begin{cases} \hat{u}_m & \text{on } \Gamma_m \in \hat{\Gamma}; \\ \hat{u}_n & \text{on } \Gamma_m \in \Gamma_S; \\ u_{\Gamma_{n,m}} - \Delta u_{m,n} & \text{on } \Gamma_{m,n}, m < n, \end{cases} \quad (A3)$$

and

$$\nabla u_m \cdot \hat{n}_i = 0, \quad \Gamma_m \in \Gamma_B, \quad (A4)$$

where $\rho(x,y) = -e[n(r) - p(r) + N_A^- - N_D^+]$ is the total charge density used in Eq. (1) of Sec. II of the main text, \hat{u}_m is the value of the potential set by the background doping and the requirement of space charge neutrality, $\hat{u}_n = 0.8 eV/k_B T$ is the near midgap surface pinned potential, and $\Delta u_{m,n} = u_{\Gamma_{n,m}} - u_{\Gamma_{m,n}}$ is the conduction band discontinuity at the heterointerface formed between regions Ω_m and Ω_n and \hat{n}_i is the unit surface normal. The charge

density term on the right-hand side of the Poisson equation will be denoted by $f_m = f(u_m)$ and is given by the Fermi-Dirac integral of order 1/2. The nonlinear nature of the Poisson equation will be discussed later.

We will use the method of variations to formulate the numerical problem. We first identify a total energy functional as

$$\begin{aligned} \Pi = & \sum_m \left\{ \int_{\Omega_m} \frac{1}{2} (-\nabla u_m)^2 d\Omega_m \right\} \\ & - \sum_m \left\{ \int_{\Omega_m} f_m u_m d\Omega_m \right\} + \sum_m \left\{ \frac{\alpha}{2} (u_{\Gamma_m} - \hat{u}_m)^2 \right\} \\ & + \sum_{m,n} \left\{ \frac{\alpha}{2} (u_{\Gamma_{m,n}} - u_{\Gamma_{n,m}} - \Delta u_{m,n})^2 \right\}. \quad (A5) \end{aligned}$$

We have used the penalty method to incorporate the linear constraint boundary conditions, where α is a large penalty factor that will be discussed later. We prefer the penalty method to the Lagrange multiplier method since it does not increase the dimensionality of the resulting linear system. We now invoke the principle of stationarity of Π with respect to the state variable u by evaluating $\delta\Pi=0$. In the calculus of variations, this is the minimization of the energy functional. We denote arbitrary variations in the state variable, which satisfy the essential boundary conditions by δ . From the minimization we obtain

$$\begin{aligned} \delta\Pi=0 = & \sum_m \left\{ \int_{\Omega_m} \nabla u_m \nabla \delta u_m d\Omega_m \right\} \\ & - \sum_m \left\{ \int_{\Omega_m} f_m \delta u_m d\Omega_m \right\} + \sum_m \{ \alpha (u_{\Gamma_m} \\ & - \hat{u}_m) \delta u_{\Gamma_m} \} + \sum_{m,n} \{ (u_{\Gamma_{m,n}} - u_{\Gamma_{n,m}} - \Delta u_{m,n}) \\ & \times (\delta u_{\Gamma_{m,n}} - \delta u_{\Gamma_{n,m}}) \}. \quad (A6) \end{aligned}$$

The state variable, u_m , is subject to the boundary conditions specified by Eqs. (A3) and (A4) above. The third term on the right-hand side is the penalty method implementation of the bulk charge neutrality conditions. These essential boundary conditions could be implemented by Dirichlet conditions with a corresponding reduction of the number of unknowns. The solution to Eq. (A5) above is unique and equivalent to the solution of the original problem.

2. Finite element formulation

The domain Ω is nonuniformly discretized with M nodal points at locations r_1, r_2, \dots, r_M , for each of which is defined an orthonormal global shape function, $\phi_i(r)$, such that $\phi_i(r_j) = \delta_{ij}$. In our calculations we use linear Chapeau shape functions. The potential u is approximated by an expansion in this basis, $u(r) \approx \sum_{i=1}^M u_i \phi_i(r) = N(r) \cdot u$, where $N(r)$ is the vector of M shape functions and u is the vector of M potentials at each nodal point. The gradient of

the potential can be likewise approximated by the expansion $\nabla u(r) \approx \sum_{i=1}^M u_i \nabla \phi_i(r) = B(r) \cdot u$, where $B(r)$ is the $2 \times M$ matrix of the gradients of the M global shape functions. The same approximations are made for the potential variations, $\delta u(r) \approx \delta u^T \cdot N^T$ and $\nabla \delta u \approx \delta u^T \cdot B^T$. Using these expansions, the variational statement of the Poisson equation can be written as

$$\begin{aligned} \delta\Pi=0 = & \sum_m \left\{ \delta u_m^T \left(\int_{\Omega_m} B^T(r) B(r) d\Omega_m \right) u_m \right. \\ & - \delta u_m^T \left(\int_{\Omega_m} N^T(r) \cdot f(u_m) d\Omega_m \right) \} \\ & + \sum_m \{ \alpha \delta u_{\Gamma_m}^T \cdot u_{\Gamma_m} \} - \sum_m \{ \alpha \delta u_{\Gamma_m}^T \hat{u}_m \} \\ & + \sum_{m,n} \left\{ \alpha \left(\delta u_{\Gamma_{m,n}} \right)^T \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{pmatrix} u_{\Gamma_{m,n}} \\ u_{\Gamma_{n,m}} \end{pmatrix} \right\} \\ & - \sum_{m,n} \left\{ \alpha \left(\delta u_{\Gamma_{m,n}} \right)^T \begin{bmatrix} 1 \\ -1 \end{bmatrix} \Delta u_{m,n} \right\}. \quad (A7) \end{aligned}$$

We now define e_{Γ_m} as a $M \times 1$ vector containing a 1 in each position corresponding to $\hat{\Gamma}_m$ and zero elsewhere. We also define the $M \times M$ global matrix $K'_{\Gamma_{m,n}}$ and the $M \times 1$ global vector $P'_{\Gamma_{m,n}}$ as follows:

$$K'_{\Gamma_{m,n}} = \begin{pmatrix} u_{\Gamma_{m,n}} & \delta u_{\Gamma_{m,n}} \\ u_{\Gamma_{n,m}} & \delta u_{\Gamma_{n,m}} \end{pmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad (A8a)$$

$$P'_{\Gamma_{m,n}} = \begin{pmatrix} u_{\Gamma_{m,n}} \\ u_{\Gamma_{n,m}} \end{pmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \quad (A8b)$$

We now define the following $M \times M$ matrix, K_m , $M \times 1$ vector $f_m(u_m)$, and $M \times 1$ vector P_m as

$$\begin{aligned} K_m = & \left\{ \int_{\Omega_m} B^T(r) B(r) d\Omega_m \right\} + \alpha e_{\Gamma_m}^T e_{\Gamma_m} \\ & + \alpha \sum_{n>m} K'_{\Gamma_{m,n}}, \quad (A9) \end{aligned}$$

$$f_m(u_m) = \left\{ \int_{\Omega_m} N^T(r) f(u_m) d\Omega_m \right\}, \quad (A10)$$

$$P_m = \alpha e_{\Gamma_m}^T \hat{u}_m + \alpha \sum_{n>m} P'_{\Gamma_{m,n}}. \quad (A11)$$

The linear constraint boundary conditions are incorporated into the stiffness matrix, K_m , by the penalty method. The penalty, α , must be several orders of magnitude larger than the largest diagonal element of the first term in Eq. (A9). Although the final solution will not depend upon the actual value of α , it must be sufficiently large to force the essential boundary conditions of the system.

We can simplify the problem by defining the global matrix K , global vector $f(u_m)$, and global vector P , such that $\sum_m \delta u_m^T K_m u_m^T = \delta u^T K u$, $\sum_m \delta u_m^T f_m(u_m) = \delta u^T P(u)$, and

$\sum_m \delta u_m^T P_m = \delta u^T P$. The finite element formulation for the Poisson equation can now be written as the following system of equations:

$$Ku = f(u) + P. \quad (A12)$$

3. Newton-Raphson formulation

The charge density term in Eq. (A10), $f(u)$, is a function of the potential, u , through the semiclassical Thomas-Fermi screening model for the carrier density, as discussed in Sec. II of the main text, $n(u) = N_c F_{1/2}(-u)$, and a similar term is obtained for the hole concentration. $N_c = 2(2\pi k_B T m_c^*/\hbar^2)^{3/2}$ is the effective conduction band density of states and $F_{1/2}$ is the Fermi-Dirac integral of order $\frac{1}{2}$. We use the Newton-Raphson iteration method to solve the nonlinear system of equations. If we let n denote the iteration count, then we can represent the solution of the $n+1$ iteration by

$$u^{n+1} = u^n + \Delta u^n, \quad (A13)$$

where Δu^n represents the change in the solution between iterations. The nonlinear function f can be approximated by a first-order Taylor series expansion:

$$f(u^{n+1}) = f(u^n) + \partial f(u^n) \Delta u^n, \quad (A14)$$

where $\partial f(u^n)$ is the partial derivative of f with respect to u at the n th iteration and is given by the Fermi-Dirac integral of order $-\frac{1}{2}$.¹⁵ We substitute this expansion for f into the system of equations as follows:

$$Ku^{n+1} = f(u^{n+1}) + P, \quad (A15)$$

$$Ku^n + K \Delta u^n = f(u^n) + \partial f(u^n) \Delta u^n + P, \quad (A16)$$

$$(K - \partial f(u^n)) \Delta u^n = f(u^n) + P - Ku^n. \quad (A17)$$

The problem has now been reduced to a system of linear equations in Δu at each iteration. We use the bulk value of the potential given by the Si-dopant concentration and the charge neutrality condition in each domain Ω_m as an initial guess, and then solve the linear system for Δu^n and update the solution u until the mean-squared difference between iterations is below a specified tolerance, typically 1×10^{-9} . A standard L/U decomposition method is used to solve the sparse linear system using an efficient skyline storage technique.

APPENDIX B: FORMULATION OF THE SCHRÖDINGER PROBLEM

1. Problem statement

The solution domain for the Schrödinger equation is described in the same manner as the Poisson domain. The quantum mechanical problem of solving Eq. (3) in Sec. II of the main text, can be stated as follows:

Find:

$$\psi = \psi_1 \cup \psi_2 \cup \dots \cup \psi_N \in H^2(\Omega), \quad (B1)$$

such that

$$-\frac{\hbar^2}{2m_c^*} \nabla^2 \psi_m(x, y) + [V(x, y) - e\phi_m(x, y)] = E\psi_m(x, y), \quad (B2)$$

and

$$\psi_m = \begin{cases} 0, & \text{on } \Gamma_m \in \hat{\Gamma} \cup \Gamma_S; \\ \psi_{\Gamma_{n,m}}, & \text{on } \Gamma_{m,n}, m < n. \end{cases} \quad (B3)$$

or

$$\nabla \psi_m \cdot \hat{n}_i = 0, \quad \text{on } \Gamma_m \in \hat{\Gamma} \cup \Gamma_S$$

and

$$\psi_m = \psi_{\Gamma_{n,m}}, \quad \text{on } \Gamma_{m,n}, m < n, \quad (B4)$$

where $V(x, y)$ includes the external potentials and $\phi_m(x, y) = u_m(x, y) \times k_B T$ is the electrostatic potential obtained from Poisson's equation and \hat{n}_i is the unit boundary normal. As indicated above, either zero-valued essential or zero-valued Neumann boundary conditions can be used to specify the decay of the wave function.

We will use the method of variations, as we did with the Poisson equation, to formulate the numerical method. We identify the total functional as

$$\begin{aligned} \Pi = \sum_m & \left[\int_{\Omega_m} \frac{1}{2} \left(-\frac{\hbar^2}{2m_c^*} \nabla \psi_m \right)^2 d\Omega_m \right] \\ & + \sum_m \left[\int_{\Omega_m} \frac{1}{2} (V_m - e\phi_m) \psi_m^2 d\Omega_m \right] \\ & - \sum_m \left[E \int_{\Omega_m} \frac{1}{2} \psi_m^2 d\Omega_m \right] + \sum_m \left[\frac{\alpha}{2} \psi_m^2 \hat{r}_m \right] \\ & + \sum_{\substack{m,n \\ m < n}} \left[\frac{\alpha}{2} (\psi_{\Gamma_{m,n}} - \psi_{\Gamma_{n,m}})^2 \right]. \end{aligned} \quad (B5)$$

The minimization of the total functional yields

$$\begin{aligned} \delta \Pi = 0 = \sum_m & \left[\int_{\Omega_m} \frac{\hbar^2}{2m_c^*} \nabla \psi_m \nabla \delta \psi_m d\Omega_m \right] \\ & + \sum_m \left[\int_{\Omega_m} (V_m - e\phi_m) \psi_m \delta \psi_m d\Omega_m \right] \\ & - \sum_m \left[E \int_{\Omega_m} \psi_m \delta \psi_m d\Omega_m \right] + \sum_m \{ \alpha \psi_m \hat{r}_m \delta \psi_m \} \\ & + \sum_{\substack{m,n \\ m < n}} \{ \alpha (\psi_{\Gamma_{m,n}} - \psi_{\Gamma_{n,m}}) (\delta \psi_{\Gamma_{m,n}} - \delta \psi_{\Gamma_{n,m}}) \}. \end{aligned} \quad (B6)$$

The state variable, ψ_m , is also subject to the boundary conditions specified by Eq. (B3) or (B4). The zero-valued

essential boundary conditions are implemented above by the penalty method, but could also be implemented by zero-valued Dirichlet conditions with a corresponding reduction of the number of unknowns. The last term in Eq. (B5) implements the continuity of the wave function across the heterointerface by the penalty method. The Schrödinger solution is therefore compatible with the Poisson solution of the electrostatic potential. The solution to Eq. (B6) above is unique and equivalent to the solution of the original problem stated in Eqs. (B1)–(B4).

$$\delta\Pi=0=\sum_m\left\{\delta\psi_m^T\left(\frac{\hbar^2}{2m_c^*}\int_{\Omega_m}\mathbf{B}^T(\mathbf{r})\mathbf{B}(\mathbf{r})d\Omega_m\right)\psi_m\right\}+\sum_m\left\{\delta\psi_m^T\left(\int_{\Omega_m}(V_m-e\phi_m(\mathbf{r}))\mathbf{N}^T(\mathbf{r})\cdot\mathbf{N}(\mathbf{r})d\Omega_m\right)\psi_m\right\}-\sum_m\left\{\delta\psi_m^T\left(E\int_{\Omega_m}\mathbf{N}^T(\mathbf{r})\cdot\mathbf{N}(\mathbf{r})d\Omega_m\right)\psi_m\right\}+\sum_m\left\{\alpha\delta\psi_{\Gamma_m}^T\cdot\psi_{\Gamma_m}\right\}+\sum_{m<n}\left\{\alpha\begin{pmatrix}\delta\psi_{\Gamma_{m,n}}^T\\\delta\psi_{\Gamma_{n,m}}^T\end{pmatrix}\begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix}\begin{pmatrix}\psi_{\Gamma_{m,n}}\\\psi_{\Gamma_{n,m}}\end{pmatrix}\right\}. \quad (\text{B7})$$

Using similar definitions as those in the Poisson development in Appendix A, this set of linear equations can be reduced to a generalized eigenvalue problem,

$$\mathbf{A}\psi=\mathbf{E}\mathbf{B}\psi. \quad (\text{B8})$$

Because we are only interested in the lowest bound states of the system, we need not solve for all of the eigenenergies and wave functions. We use the subspace iteration method with efficient skyline storage sparse matrix methods to solve for the lowest ten bound states.

¹J. M. Ballingall and C. E. C. Wood, *Appl. Phys. Lett.* **41**, 947 (1982).

²W. I. Wang, E. E. Mendez, T. S. Kuan, and L. Esaki, *Appl. Phys. Lett.* **47**, 826 (1985).

³S. Subbanna, H. Kroemer, and J. L. Merz, *J. Appl. Phys.* **59**, 488 (1986).

⁴D. L. Miller, *Appl. Phys. Lett.* **47**, 1309 (1985).

⁵J. Ebner, J. E. Lary, G. W. Eliason, and T. K. Plant, *Proceedings of the 20th European Solid State Device Research Conference, ESSDERC 90*, edited by W. Eccleston and P. J. Rossner (Hilger, Bristol, 1990), pp. 401–404.

⁶H. K. Harbury, W. Porod, and S. M. Goodnick, presented at The 19th

2. Finite element formulation

The domain, Ω , for the Schrödinger problem uses the same generalized discretization scheme developed for the Poisson problem and is schematically shown in Fig. 15. Using linear Chapeau basis functions we expand $\psi(\mathbf{r})$ and $\delta\psi(\mathbf{r})$ in the finite element approximations: $\psi(\mathbf{r})\approx\mathbf{N}(\mathbf{r})\cdot\psi$, $\nabla\psi(\mathbf{r})\approx\mathbf{B}(\mathbf{r})\cdot\psi$, $\delta\psi(\mathbf{r})\approx\delta\psi^T\cdot\mathbf{N}(\mathbf{r})^T$, and $\nabla\delta\psi(\mathbf{r})\approx\delta\psi^T\cdot\mathbf{B}(\mathbf{r})^T$. With these expansions, the variational formulation of the problem becomes

International Conference on the Physics and Chemistry of Semiconductor Interfaces, Death Valley, CA, January 1992.

⁷W. Porod, H. K. Harbury, and S. M. Goodnick, *Appl. Phys. Lett.* **61**, 1823 (1992).

⁸C. C. Dean and M. Pepper, *J. Phys. C* **15**, L1287 (1982).

⁹A. B. Fowler, A. Hartstein, and R. A. Webb, *Phys. Rev. Lett.* **48**, 196 (1982).

¹⁰B. I. Miller, A. Shahar, U. Koren, and P. J. Corvini, *Appl. Phys. Lett.* **54**, 188 (1989).

¹¹E. Kapon, D. M. Hwang, and R. Bhat, *Phys. Rev. Lett.* **63**, 430 (1989); E. Kapon, S. Simhony, R. Bhat, and D. M. Hwang, *Appl. Phys. Lett.* **55**, 2715 (1989).

¹²E. Colas, S. Simhony, E. Kapon, R. Bhat, D. M. Hwang, and P. S. D. Lin, *Appl. Phys. Lett.* **57**, 914 (1990).

¹³H. K. Harbury and W. Porod, *J. Vac. Sci. Technol. B* **8**, 923 (1990).

¹⁴*Nanostructure Physics and Fabrication*, edited by M. A. Reed and W. P. Kirk (Academic, San Diego, CA, 1989).

¹⁵C. M. Wolfe, N. Holonyak, Jr., and G. E. Stillman, *Physical Properties of Semiconductors* (Prentice-Hall, Englewood Cliffs, NJ, 1989).

¹⁶Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology.

¹⁷E. H. Rhoderick and R. H. Williams, *Metal-Semiconductor Contacts* (Oxford University Press, Oxford, 1988).

Bistable saturation in coupled quantum-dot cells

P. Douglas Tougaw, Craig S. Lent, and Wolfgang Porod

Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

(Received 29 April 1993; accepted for publication 18 May 1993)

Model quantum dot cells are investigated as potential building blocks for quantum cellular automata architectures. Each cell holds a few electrons and interacts Coulombically with nearby cells. In acceptable cell designs, the charge density tends to align along one of two cell axes. Thus, a cell "polarization," which can be used to encode binary information, is defined. The polarization of a cell is affected in a very nonlinear manner by the polarization of its neighbors. This interaction is quantified by calculating a cell-cell response function. Effects of nonzero temperature on the response of a model cell are investigated. The effects of multiple neighbors on a cell are examined and programmable logic gate structures based on these ideas are discussed.

I. INTRODUCTION

For many years, the size of microelectronic devices has been shrinking, and this has led to faster, denser circuits. Despite these improvements, the basic computing paradigm has remained virtually unchanged because device operation has been largely unaltered, apart from rescaling. There is now much interest in extremely dense device arrays forming locally interconnected architectures like cellular automata (CA)¹ and cellular neural networks.² Such architectures could lead to changes in device structure of a less evolutionary and more revolutionary nature.

At the same time, many researchers have been investigating ways to use quantum structures as electronic devices. In the course of such research, a great deal has been learned about the behavior of electrons in very small structures. Because of the size of the structures involved, an outstanding difficulty is providing a scheme in which one of these quantum devices, which typically carry nanoamperes of current, could be used to drive several other similar devices. In addition, the capacitance of the wires needed to interconnect such structures would tend to dominate their behavior. Therefore, locally connected architectures like CA's may be an attractive paradigm for implementing quantum device architectures.³

CA architectures composed of nanometer-scaled quantum devices that are coupled through the Coulomb interaction (no current flows between devices) have been proposed by the authors elsewhere.⁴⁻⁶ We call such architectures quantum cellular automata (QCA). The QCA contains an array of quantum-dot cells that are connected locally by the interactions of the electrons contained within them. The quantum state of each multidot cell encodes the "logical" state of that cell. For this reason, each cell should ideally have exactly two stable states, since this will allow direct encoding of binary information.⁷ Such two-state cells also need to exhibit bistable saturation to ensure that noise or small geometric variations do not overwhelm the signal.

To function as a CA, the state of each cell should be dependent on the states of its neighbors. In this paper, we compare the cell-cell coupling and bistable saturation of

several different quantum cell designs that might form the basis of quantum cellular automata. All these designs have certain characteristics in common: a few (typically four or five) quantum dots connected by coupling coefficients and populated by a total of one to three electrons. In these cells the required interaction between neighbors is caused by the mutual Coulombic repulsion of the electrons contained in the cells. We use a very simple model of each cell, neglecting details relating to exactly how the quantum dot structures are realized, but focusing on the charge distribution among the dots and the Coulomb coupling between cells. We define a cell-cell response function that characterizes the interaction between neighboring cells.

In the next section we will introduce the theoretical model of the "standard cell," on which much of the work of Refs. 4-6 is based. It is the most thoroughly investigated cell design because it displays strong bistable saturation. We discuss the model Hamiltonian used for the cell, the method used to calculate the cell-cell response function, and the effects of nonzero temperature. In Sec. III we compare various other cell designs. Among these are different geometric arrangements of the quantum dots, one and three electron cells, and continuous quantum dashes. In Sec. IV, we extend our results to include the effects of multiple neighbors on a cell. We show that such effects in a system with three nearest neighbors can be thought of as majority voting logic. We show how this behavior can be used to implement programmable logic gates, and then show other possible implementations for dedicated AND and OR gates. A discussion and conclusion follow in Sec. V.

II. A MODEL QUANTUM CELL

The model "standard cell" design, shown schematically in Fig. 1(a), consists of five quantum dots located at the corners and the center of a square. Tunneling occurs between the central site and all four of the outer sites (near-neighbor tunneling), and to a lesser degree between neighboring outer sites (next-near-neighbor tunneling). It is assumed that the potential barriers between cells are high enough to completely suppress intercellular tunneling.

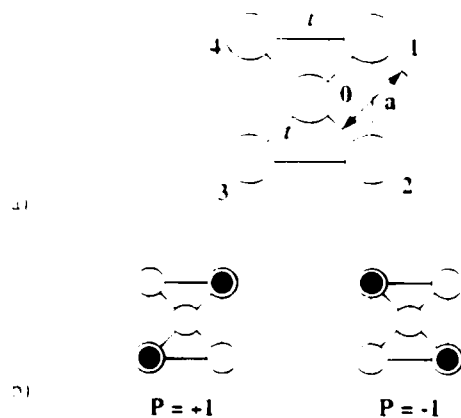


FIG. 1. Schematic of the basic five-site cell. (a) The geometry of the cell. The tunneling energy between the middle site and an outer site is designated by t , while t' is the tunneling energy between two outer sites. (b) Coulombic repulsion causes the electrons to occupy antipodal sites within the cell. These two bistable states result in cell polarizations of $P = +1$ and $P = -1$ [see Eq. (1)].

The cell is occupied by a total of two electrons hopping among the five sites; these electrons tend to occupy antipodal outer sites within the cell due to their mutual electrostatic repulsion [see Fig. 1(b)].

We will show that these two stable states are degenerate in an isolated cell, but an electrostatic perturbation in the cell's environment (such as that caused by neighboring cells) splits the degeneracy and causes one of these configurations to become the cell ground state. Altering the perturbation causes the cell to switch between the states in an abrupt and nonlinear manner. This very desirable bistable saturation behavior is due to a combination of quantum confinement, Coulombic repulsion, and the discreteness of electronic charge.

A. Cell polarization

Since Coulomb repulsion causes the electrons to occupy antipodal sites, the ground state charge density may have the electrons aligned along one of the two diagonal axes shown in Fig. 1(b). We therefore define the cell polarization, a quantity that measures the extent to which the charge density is aligned along one of these axes. The polarization is defined as

$$P \equiv \frac{(\rho_1 + \rho_3) - (\rho_2 + \rho_4)}{\rho_0 + \rho_1 + \rho_2 + \rho_3 + \rho_4}, \quad (1)$$

where ρ_i denotes the electron probability density at site i . As in Fig. 1(b), electrons completely localized on sites 1 and 3 will result in $P = 1$, while electrons on sites 2 and 4 yield $P = -1$. An isolated cell would have a ground state that is a linear combination of these two states, and would therefore have a net polarization of zero.⁸

B. Model cell Hamiltonian

We employ a simple model of the quantum cell that uses a tight-binding Hubbard-type Hamiltonian. We rep-

resent the quantum dots as sites, ignoring any degrees of freedom internal to the dot. The Hamiltonian for a single isolated cell can be written as

$$H_0^{\text{cell}} = \sum_{i,\sigma} E_0 n_{i,\sigma} + \sum_{i,j} t_{ij} (a_{i,\sigma}^\dagger a_{j,\sigma} + a_{j,\sigma}^\dagger a_{i,\sigma}) - \sum_i E_Q n_{i,\uparrow} n_{i,\downarrow} + \sum_{i,j,\sigma,\sigma'} V_Q \frac{n_{i,\sigma} n_{j,\sigma'}}{|\mathbf{R}_i - \mathbf{R}_j|}. \quad (2)$$

Here $a_{i,\sigma}$ is the annihilation operator that destroys a particle at site i ($i=0,1,2,3,4$) with spin σ . The number operator for site i and spin σ is represented by $n_{i,\sigma}$. E_0 is the on-site energy for each dot, t_{ij} is the energy associated with tunneling between dots i and j , and E_Q is the on-site charging energy (the purely Coulombic cost for two electrons of opposite spin to occupy the same dot). The last term in (2) represents the Coulombic potential energy due to electrons on the i th and j th sites at positions \mathbf{R}_i and \mathbf{R}_j . V_Q is an electrostatic parameter fixed by fundamental constants and the dielectric constant of the material used to form the cells.

For the cell described above we use values of the parameters in the Hamiltonian based on a simple, experimentally reasonable model. We take E_0 to be the ground state energy of a circular quantum dot of diameter 10 nm holding an electron with effective mass $m^* = 0.067 m_0$. The near-neighbor distance between dot centers, a , is taken to be 20 nm. The Coulomb coupling strength, V_Q , is calculated for a material with a dielectric constant of 10, and E_Q is taken to be $V_Q/(D/3)$.⁹ The coupling energy between an outer dot and the central dot is $t \equiv t_{0,i} = 0.3$ meV ($i=1,2,3,4$), and the next-near-neighbor coupling connecting the outer dots, t' , is taken to be $t/10$ (consistent with one-dimensional calculations for a barrier height of 150 meV).¹⁰

The spins of the two electrons in the cell can be either parallel or antiparallel. We consider here the case of electrons with antiparallel spins, since that is the ground state of the cell. Calculations with electrons having parallel spins yield qualitatively very similar results.

To maintain charge neutrality in the cell, a fixed positive charge $\bar{\rho}$, corresponding to a charge of $(2/5)e$, is also assumed on each site. In a single isolated cell this just renormalizes E_0 , but the fixed charge is important when simulating systems with more than one cell. If each cell had a net negative charge, then electrons near the edges of a group of cells would respond mostly to the net negative charge of the other cells. In a semiconductor realization, the fixed positive charge would likely be provided by ionized donor impurities and charge on the surface of metal gates.

The interaction of a cell with its electrostatic environment (including neighboring cells) is contained in a second Hamiltonian term, which we write as $H_{\text{inter}}^{\text{cell}}$. We solve the time-independent Schrödinger equation for the n th eigenstate of the cell, $|\Psi_n\rangle$, under the influence of the neighboring cells:

$$(H_0^{\text{cell}} + H_{\text{inter}}^{\text{cell}}) |\Psi_n\rangle = E_n |\Psi_n\rangle. \quad (3)$$

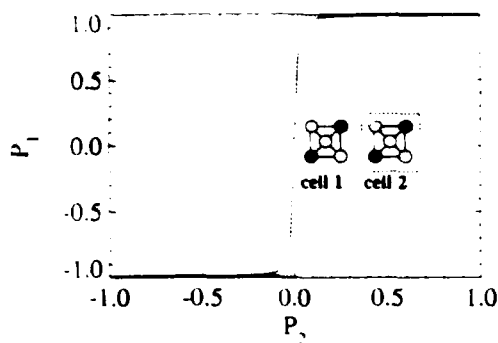


FIG. 2. Cell-cell response function for the basic five-site cells shown in the inset. This shows the polarization P_1 induced in cell 1 by the fixed polarization of its neighbor, P_2 . The solid line corresponds to antiparallel spins, and the dotted line to parallel spins. The two are nearly degenerate, especially for significantly large values of P_2 .

The total Hamiltonian is diagonalized directly in the basis of two-particle site kets. We calculate the single-particle density, ρ_i , from the two-particle ground-state wave function $|\Psi_0\rangle$ by finding the expectation value of the number operator for site i :

$$\rho_i = \sum_{\sigma} \langle \Psi_0 | n_{i,\sigma} | \Psi_0 \rangle. \quad (4)$$

We can then use these densities to calculate the cell polarization P as in Eq. (1).

C. Calculating the cell-cell response function

To be useful in cellular automata-type architectures, the state of a cell must be strongly influenced by the states of neighboring cells. To demonstrate how one of these cells is influenced by the state of its neighbor, consider the two cells shown in the inset to Fig. 2. The cell centers are separated by a distance of $3a = 60$ nm. We assume cell 2 has a given polarization P_2 and that the electron density on the central site is negligible. This means that the charge density is completely determined by the cell polarization. For the corresponding electron density on each site of cell 2, we calculate the electrostatic potential at each site of cell 1. This additional potential energy is then included in the Hamiltonian for cell 1. Thus the perturbing Hamiltonian component is

$$H_{\text{inter}}^{\text{cell}} = H_1^{\text{cell}} = \sum_{i \in \text{cell 1}, \sigma} V_i^1 n_{i,\sigma}, \quad (5)$$

where

$$V_i^m = \sum_{k \neq m, j} V_Q \frac{(\rho_j^k - \bar{\rho})}{|\mathbf{R}_{k,j} - \mathbf{R}_{m,i}|} \quad (6)$$

is the potential at site i in cell m due to the charges in all other cells. We denote the position of site j in cell k as $\mathbf{R}_{k,j}$, and the electron density at site j in cell k as ρ_j^k . The total Hamiltonian for cell 1 is then

$$H^{\text{cell}} = H_0^{\text{cell}} + H_1^{\text{cell}}. \quad (7)$$

The two-electron time-independent Schrödinger equation is solved using this Hamiltonian for a series of values of P_2

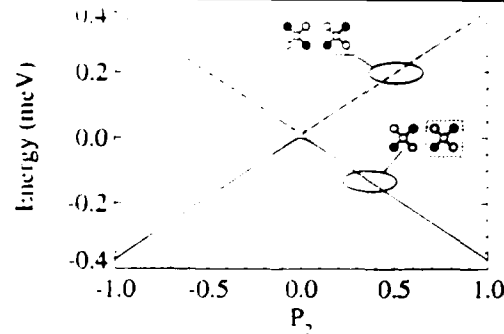


FIG. 3. The lowest four eigenstate energies of cell 1 induced by the polarization of cell 2. The insets show that the lowest two energy states always correspond to the same polarization direction, as in the driver. Slight exchange splitting effects between spatially symmetric and spatially antisymmetric states breaks the fourfold degeneracy for very small values of P_2 .

in the range $[-1, +1]$. The ground state polarization of cell 1, P_1 , is then computed for each value of P_2 , as described in the previous section. Thus, we can plot the induced polarization of cell 1 as a function of the polarization of cell 2. This function $P_1(P_2)$, which we call the cell-cell response function, is one measure of how well a cell will operate in a quantum cellular automaton architecture.

Figure 2 shows the cell-cell response function for the standard cell. The highly nonlinear nature of the response indicates that a small polarization in cell 2 causes a very strong polarization in its neighbor, cell 1. The figure also shows that the polarization of cell 1 saturates very quickly to a value of $+1$ or -1 . This bistable saturation is the basis of the quantum cellular automata, since it means that we can encode bit information using the cell polarization. We assign the bit value of 1 to the $P = +1$ state and the bit value 0 to the $P = -1$ state. Since the cell is almost always in a highly polarized state ($|P| \approx 1$), the state of the cell will be indeterminate only if the electrostatic environment due to other cells is perfectly symmetric.

Figure 3 shows the lowest four eigenenergies of cell 1 as a function of the polarization of cell 2. This shows that the perturbation due to the polarization of cell 2 quickly separates the states of opposite polarization. For a completely polarized standard cell, the excitation energy from the ground state to the first excited state with opposite polarization is about 0.8 meV. This corresponds to a temperature of about 9 K.

The abruptness of the cell-cell response function depends on the ratio of the tunneling energy, t in Eq. (2), to the Coulomb energy for electrons on neighboring sites. The magnitude of the tunneling energy depends exponentially on both the distance between dots and the height of the potential barrier between them. Figure 4 shows the cell-cell response function for different values of the tunneling energy t . The switching becomes more abrupt as t decreases. Of course, if t goes to zero, the tunneling would be completely suppressed and no switching would occur. Extremely small values of t would similarly slow the switch-

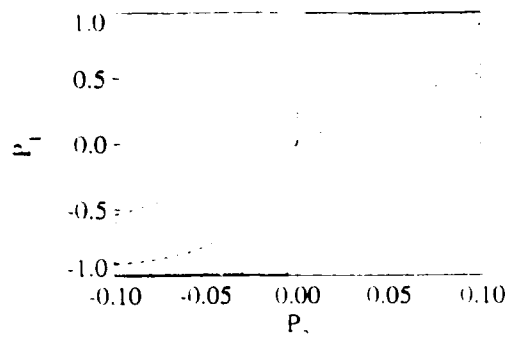


FIG. 4. The cell-cell response function of the basic five-site cell for various values of the parameter t in Eq. (2) (after Ref. 4). The induced polarization in cell 1 is plotted as a function of the polarization of its neighbor, cell 2. The curves correspond to $t=0.2$ meV (solid line), $t=0.3$ meV (dotted line), $t=0.5$ meV (dashed line), and $t=0.7$ meV (dot-dashed line). Note the horizontal axis only shows P_2 in the range $[-0.1, +0.1]$.

ing time. For $t=0.3$ meV, the standard cell value, we estimate the tunneling time as $\hbar/t=2$ ps.

D. Nonzero temperature cell-cell response

We extend the calculation of the cell-cell response function to nonzero temperatures by calculating the thermal expectation value of the electron density at each site of the cell,

$$\rho_i = \langle \hat{n}_i \rangle = \frac{\sum_{\sigma} \sum_n \langle \Psi_n | \hat{n}_{i,\sigma} | \Psi_n \rangle \cdot e^{-E_n/(k_B T)}}{\sum_n e^{-E_n/(k_B T)}}. \quad (8)$$

Evaluating the thermal average requires knowledge of the excited states of the cell as well as the ground states. Using the results of Eq. (8), the polarization of the cell can be calculated as before using Eq. (1). The results of such a calculation for the standard cell (with the next-near-neighbor coupling $t'=0$) are shown in Fig. 5. The curve for $T=0$ is the same as in Fig. 2. The nonlinearity of the response degrades as the temperature increases. For tem-

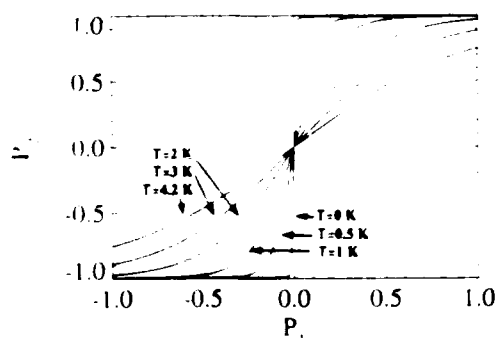


FIG. 5. The cell-cell response function for the standard cell with $t'=0$ at various temperatures. The response degrades as temperature increases. Above 4.2 K, the response would be unacceptable for use in a QCA. This maximum operating temperature is highly dependent on the physical size of the cell; molecule-sized cells would behave in a satisfactory manner up to room temperature.

peratures up to 4.2 K, the response is good, but for higher temperatures it would probably be unacceptable for use in a QCA.

Note that this maximum operating temperature will increase as the size of the cell decreases. Thus, although a 20 nm design rule requires cryogenic temperatures for satisfactory operation, the design scales to smaller sizes easily and a much smaller, possibly macromolecular, implementation¹¹ would work at room temperature.

III. ALTERNATIVE QUANTUM CELLS

While the cell described above has demonstrated an excellent cell-cell response, there is no reason to discount other possible cell designs. Slight modifications to this cell give rise to a family of similar cells whose behavior can provide insight into the nature of the system.

In looking for other model cells, there are several approaches we can take. The most obvious of these is to alter the number of sites and their geometric arrangement. Alternatively, the cell occupation can be altered. Finally, tunneling between the intracellular sites can be increased, decreased, or effectively eliminated by varying the potential barriers between the sites.

A. Four quantum cells

In this section we will investigate the cell-cell response of four different quantum cells. While these four are representative of the sort of cells one might consider, they in no way exhaust the study of new cell designs.

The first cell, included mainly as a standard by which to judge the others, is the original cell described above with $t=0.3$ meV. This will be called cell A. Next will be the same cell with no tunneling between the outer neighbors ($t'=0$). We will refer to this as cell B. Cell C omits the presence of the central site and allows tunneling only between the four outer sites. Finally, cell D inhibits tunneling even further, allowing it only between sites 1 and 2 and between sites 3 and 4. Schematic diagrams of these four cell designs are shown in Fig. 6(a).

In a semiconductor realization of these cells, the minimum spacing between nearest-neighbor sites will be limited by the fabrication technology. For this reason, these cells are designed with a constant 20 nm design rule (the nearest neighbors in each cell are 20 nm apart). While this makes cells C and D smaller, it is the most physically reasonable way to compare their operation. The spacing between interacting cells is set at three times the near-neighbor dot spacing.

It is possible to consider cell B as an approximation to cell A that neglects tunneling between outer neighbors. In reality there will always be a certain amount of tunneling between outer sites, but this tunneling can be made arbitrarily small by selectively increasing the potential barriers between the outer sites. The same increase in potential barriers would be needed to suppress horizontal tunneling in cell D.

Figure 6(b) shows the cell-cell response functions for these four cells. This figure shows that cells A and B have very similar responses, and both are superior to cells C and

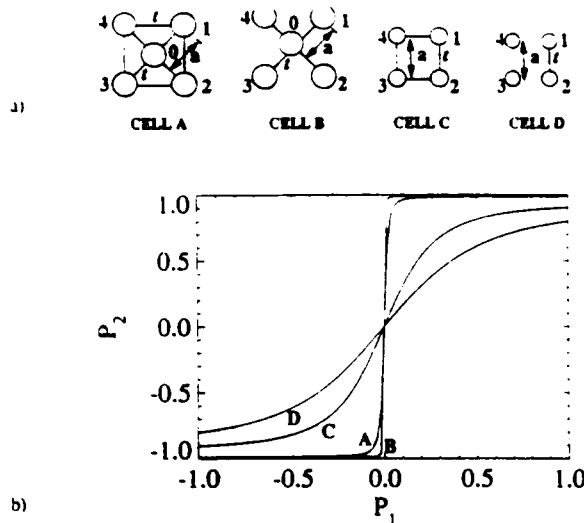


FIG. 6. Four geometric variations on the simple model quantum cell. (a) Schematic diagrams of the four cells. Cells C and D occupy less area, but all four cells are drawn with the same minimum spacing between neighbors. Cells B and D will require potential variation between the sites to selectively inhibit tunneling. (b) The cell-cell response function for each of these cell designs. Cell B has the best response, but the improvement over A is small.

D. Thus, elimination of the central site as in cells C and D degrades the response. This leads us back to the five-site cell we originally considered. Since the complete suppression of next-near-neighbor coupling as in cell B might introduce additional fabrication difficulty with little improvement in the cell response, cell A may be the most practical of these four cell designs.

B. One- and three-electron cells

As an alternative to changing the geometry of the cell, we can also alter the electron occupancy. Figure 7 shows the cell-cell response function for cell A occupied by a single electron, and Fig. 8 shows the response for the same cell with three electrons (two parallel spins, one antiparallel). These nearly linear response functions never become

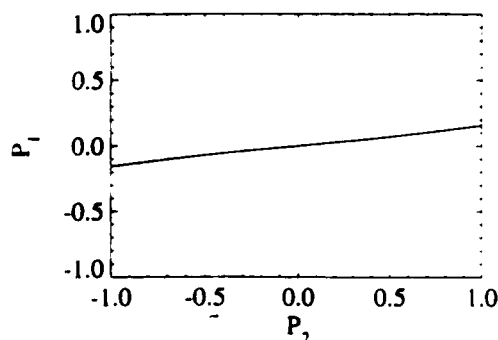


FIG. 7. The cell-cell response function for the basic five-site cell occupied by a single electron. The weak response indicates that such a cell is unsuitable as the basis of a QCA.

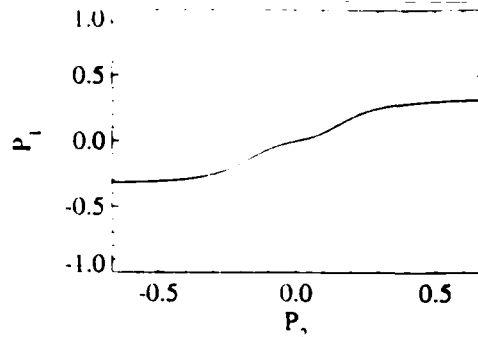


FIG. 8. Cell-cell response function for the basic five-site cell occupied by three electrons. Such a cell is also unsuitable as the basis of a QCA.

strongly polarized, even for fully polarized neighbors. This indicates that such cells would perform very poorly as the basis of a quantum cellular automaton.¹²

C. Quantum dashes and double wells

Proposals have been made for one-electron "quantum dash" cells that appear qualitatively similar to the cells we have discussed here.¹³ In this section we investigate the cell-cell response function of single-electron quantum dashes and compare this to a very similar double quantum well to show how important the discreteness of electronic charge is to the nonlinearity of the response functions seen in the previous sections.

Since these cells are of a more spatially continuous nature than cells previously considered, the site representation is no longer useful. Each cell will be modeled as a one-dimensional hard-walled square well of width 30 nm. The two cells are separated by a distance of 20 nm. These dimensions are similar to those of the cells described above. We use the finite element method to solve the single-electron time-independent Schrödinger equation for each one-electron cell. The geometry used to calculate the cell-cell response function is shown schematically in Fig. 9(a).

Since these cells have only a single axis along which to distribute the electronic charge, a new definition of polarization must be introduced. The new definition, which takes into account the continuous nature of the probability density, is

$$P = \frac{\int_{-L/2}^0 \rho(x) dx - \int_0^{L/2} \rho(x) dx}{\int_0^L \rho(x) dx} \quad (9)$$

Because of its continuous nature, the charge density in the "driver" cell is no longer uniquely determined by specifying the cell polarization. We therefore fix the charge density to be constant in each half of the driver cell.

The cell-cell response function calculated for such a system is shown in Fig. 9(b). As this figure shows, the response is quite linear, and cell 1 is virtually unpolarized, even for a fully polarized neighbor. The electron probability density as a function of position for cell 1 with a fully polarized neighbor ($P_2=1$) is shown in Fig. 9(c). The probability density is nearly symmetric about the center of the cell as we would expect for such a small polarization.

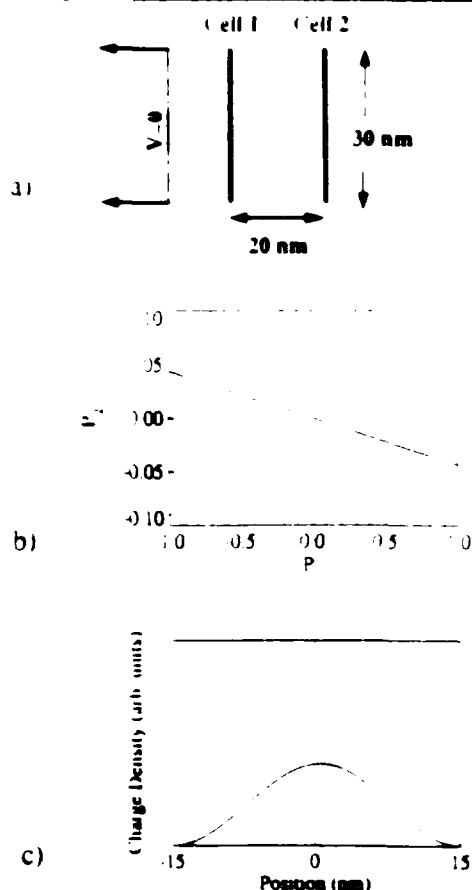


FIG. 9. The "quantum dash" as a QCA cell. (a) A schematic diagram of the cellular arrangement. The length and spacing is similar to that of the basic five-site cell in Fig. 1(a). Each cell is modeled as a one-dimensional infinite square well. The cell-cell response is shown in (b). Note that the vertical axis only shows P_2 over the range $[-0.1, +0.1]$. (c) The one-dimensional charge density in cell 2 for a fully polarized neighbor ($P_1 = 1$). The nearly symmetric charge density yields a very low polarization.

A related cell, the double well, is shown schematically in Fig. 10(a). It is a quantum well of the same dimensions as in Fig. 9(a), but the potential in the middle third of the well has been raised by 150 meV. This cell is also very similar to half of cell D from the last section, so we would expect its response to be much better than that of the simple quantum dash.

The calculated response, shown in Fig. 10(b), is indeed much better than that of Fig. 9(b). Its nonlinearity and saturation properties are very similar to those of cells C and D in Fig. 6(b). This response shows that one-electron cells can be used to provide the required nonlinear response, but it is also possible to view each pair of these cells as a single two-electron cell, which becomes geometrically very similar to cell D of Fig. 6(a).

The fact that such a seemingly small change in the nature of the cell should cause such a profound change in the cell-cell response function is linked to the fact that electron charge is discrete in regions surrounded by high potential barriers. That is to say, the expectation value of the number operator approaches an integer value as the region becomes more and more isolated by the potential

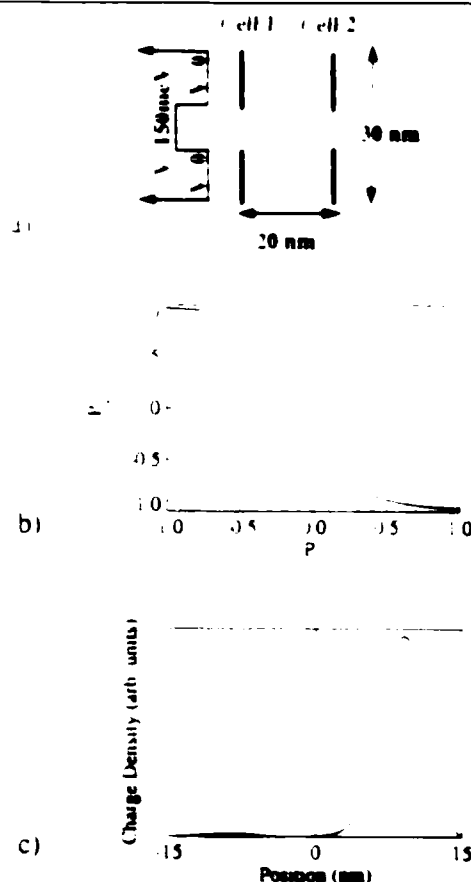


FIG. 10. The double well as a QCA cell. (a) A schematic diagram of the cellular arrangement. The total dimensions are identical to those of Fig. 8. The middle third of each cell contains a 150 meV barrier to isolate the top and bottom of the cell. (b) The cell-cell response function for such an arrangement. Note that the vertical axis now shows P_2 over the range $[-1.0, +1.0]$. (c) The one-dimensional charge density in cell 2 for a fully polarized neighbor ($P_1 = 1$). The highly asymmetric charge density results in a cell that is almost completely polarized.

barriers surrounding it.¹⁴ Therefore, almost the entire wavefunction will become localized in one-half of the cell if a small asymmetry in the electrostatic environment is introduced. This fact is demonstrated in Fig. 10(c), which shows nearly all the charge density on the right half of the cell. Since there is no barrier in the middle of the quantum dash to isolate the top and bottom of the cell, no such localization behavior is seen there, and the charge density is always nearly symmetric about the center of the well.

IV. MULTIPLE NEIGHBOR INTERACTIONS

Thus far, we have only considered the interaction between a cell and a single neighboring cell. The natural extension of this is to investigate the effects of multiple neighbors on the state of a cell. Since this implies considering a system that contains several cells and therefore several electrons, we cannot use the direct solution method described earlier for treating a single cell. For the analysis of such systems, we treat the physics within each cell as before, including exchange and correlation effects exactly. The intercellular interaction is treated self-consistently us-

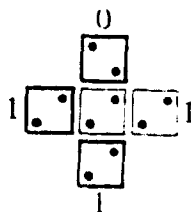


FIG. 11. Majority voting logic. The states of the center and right cells are always the same as the majority of the three fixed neighbors. The cells with heavy borders have fixed charge densities. These are not schematic diagrams; they are the actual results of the ICHA solution of the ground state charge densities in this system. The diameter of each dot is proportional to the charge density on that site.

ing a Hartree approximation. This method, called the Inter-cellular Hartree Approximation (ICHA) is detailed in Refs. 5 and 6.

Figure 11 shows an arrangement of standard cells, such that one cell has multiple neighbors. The charge densities of the cells on the top, left, and bottom are fixed, while those of the middle and right cells are free to react to the fixed charge. In an actual QCA, the states of the neighbors would not be fixed; they would be driven by the results of previous calculations or come from inputs at the edge of the QCA.

In the specific state shown in Fig. 11, two of the fixed neighbors are in the "one" state, and the other is in the "zero" state. When the ICHA is used to determine the ground state of this system, we find that the states of the center and right cells match the state of the majority of the fixed neighbors. We refer to this feature of the cell behavior, which is true for all combinations of the three inputs, as *majority voting logic*. Note that Figs. 11–13 are not schematic, but plots of the self-consistent electron density on each site. The radius of each dot is proportional to the single-electron density at that site.

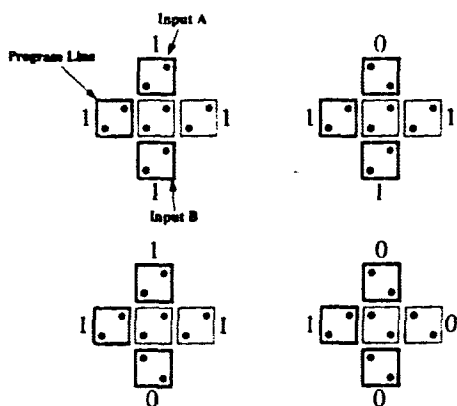


FIG. 12. The programmable AND/OR gate. The program line is set to one in each system, so the gate is displaying OR logic. All four combinations of the nonprogram line inputs are shown. The cells with heavy borders have fixed charge densities. Any one of the three inputs could be the program line; the left cell is not special. These are not schematic diagrams; they are the actual results of the ICHA solution of the ground state charge densities in each system. The diameter of each dot is proportional to the charge density on that site.

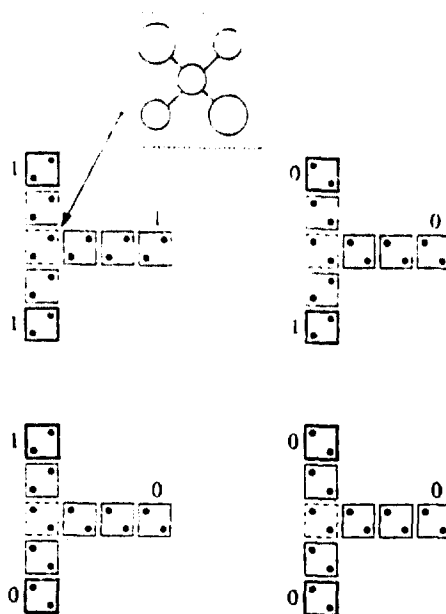


FIG. 13. The nonprogrammable AND gate. All four combinations of the inputs are shown. The cells with heavy borders have fixed charge densities, while those with dotted borders are geometrically biased toward zero as shown in the inset. The bias is sufficient to decrease the on-site energy of the affected sites by 1%. Note that the output only equals one if both of the inputs are also one. These are not schematic diagrams; they are the actual ICHA results of the ground state charge densities in each system. The diameter of each dot is proportional to the charge density on that site.

While majority voting logic behavior is valuable by itself, its potential functionality is shown by a particular interpretation of the three inputs. In Fig. 12, we have singled out one of the three and called it the program signal. Note that any one of the three neighbors could serve as the program signal, but the one case we are showing (with the program line coming in from the left) is sufficient for illustration purposes. The four systems shown include all possible combinations of signals on the two nonprogram lines. Since all four systems in Fig. 12 show the program line in the "one" state, the central cell can only be zero if the other two inputs are both zero. Thus the system realizes the truth table of the OR operation. Likewise, if the program signal is zero, the result is zero unless both of the other inputs are one. This is a realization of the AND operation.

By interpreting any one of the inputs as a program line, we have implemented a programmable AND/OR gate. The nature of this gate (AND versus OR) is defined by the state of the program line, and the other two inputs are applied to the gate thus defined.

The fact that the rightmost cell always matches the central cell means that the result of this calculation can be propagated away from the gate, down a QCA "wire,"¹⁵ and eventually serve as the input to subsequent gates. It is necessary to distinguish between driving neighbors and driven neighbors in this system. Since the rightmost cell is free to react to the states of its neighbors, it is a driven neighbor. Its state will always match that of the central cell, so only the three driving neighbors are involved in the

majority voting. Of course, once the signal is propagated away from this gate, the outgoing cells are being driven and can be used as driving neighbors for subsequent gates.

A dedicated, nonprogrammable implementation of the AND gate is shown in Fig. 13. This system has only two driving inputs; there is no program signal. The role previously played by the program signal, biasing the central cell so it can only be in the one state if both of its neighbors are, is performed by slightly enlarging the two quantum dots on sites 1 and 3 in the central cell.¹⁰ This means that the ground state of the isolated cell is no longer an unpolarized state; the cell is biased toward the zero state and can only be persuaded to enter the one state if both of its driving neighbors are one. Again, the signal propagates away to the right and can be used to drive subsequent gates. A dedicated OR gate can similarly be implemented by enlarging sites 2 and 4, biasing the cell toward the one state. It will only be in the zero state if both of its driving neighbors are also zeros.

V. DISCUSSION

With the above results, we have demonstrated several quantum-dot cells suitable for implementing a quantum cellular automata architecture. Examination of the cell-cell response function shows that for appropriate cell designs, the state of a cell is influenced very strongly by the state of its neighbors. The highly nonlinear response of the cell suggests that a signal that has become degraded by noise will be restored to full polarization by subsequent cells in the array.¹⁵ In this way, the bistable saturation of the quantum cell is analogous to the gain in a conventional digital device.

We have assumed throughout that the many electron system is in its ground state. In general, a system will start in the ground state and then be driven into an excited state by externally changing the states of input cells near the edge of a QCA array. Inelastic processes, which are usually very detrimental to the operation of quantum devices, then drive the system back to a new ground state corresponding to the new inputs. The details of the temporal evolution of the many-electron system as it relaxes to its ground state are very complicated. In the QCA scheme, we rely on the properties of the system ground state and not the details of the relaxation process for doing the computation. This idea of "computing with the ground state" and the related concept of "edge-driven" systems are discussed more thoroughly in Ref. 5.

The behavior of lines of these cells, the most basic (and important) components of a quantum cellular automaton, is discussed in Ref. 6. The results show an excellent example of the restoration of full signal strength after degradation by noise. In addition, it shows that the particular set of parameters we chose in Sec. II B is not critical; there is a wide range of parameter values for which the cells transmit information from one cell to another.

Clearly, fabrication of these devices presents a major challenge in the realization of QCA devices, but semiconductor realizations of such systems using new nanolithographic techniques should be possible. It is also possible

that future realizations of these cells will be on a macromolecular basis. Another challenge, sensing the presence or absence of a single electron without disturbing the system, necessary for reading the output state of a QCA device, has been successfully addressed.¹⁷

In conclusion, we have explored the interaction of neighboring quantum-dot cells. We have defined the cell-cell response function, which characterizes the nonlinearity of the coupling between cells, and thus determines suitability of a particular cell design for quantum cellular automata implementations. Several cell designs that exhibit the required nonlinear response and bistable saturation have been examined. Temperature effects degrade cell performance, but analysis in this simple model suggests that operation at 4.2 K should be within the reach of semiconductor implementations. When a cell has several neighbors, its state is determined by the state of the majority of the neighboring cells. This majority voting logic makes possible the construction of programmable AND/OR logic gates as well as dedicated AND and OR gates.

ACKNOWLEDGMENTS

We gratefully acknowledge stimulating conversations with Gary H. Bernstein of the Notre Dame Nanoelectronic Group. This work was supported in part by the Air Force Office of Scientific Research and the Office of Naval Research. This material is based in part upon work supported under a National Science Foundation Graduate Fellowship.

¹T. Toffoli and N. Margolus, *Cellular Automata Machines: A New Environment for Modeling* (MIT Press, Cambridge, MA, 1987).

²L. D. Chua and L. Yang, *IEEE Trans. Circuits Syst.* **35**, 1257, 1273 (1988).

³R. T. Bate, *Bull. Am. Phys. Soc.* **22**, 407 (1977); J. N. Randall, M. A. Reed, and G. A. Frazier, *J. Vac. Sci. Technol. B* **7**, 1398 (1989); D. K. Ferry, L. A. Akers, and E. W. Greeneich, *Ultra Large Scale-Interconnected Microelectronics* (Prentice-Hall, Englewood Cliffs, NJ, 1988); J. N. Randall, A. C. Seabaugh, Y.-C. Kao, J. H. Luscombe, and B. L. Newell, *J. Vac. Sci. Technol. B* **9**, 2893 (1991).

⁴C. S. Lent, P. D. Tougaw, and W. Porod, *Appl. Phys. Lett.* **62**, 714 (1993).

⁵C. S. Lent, P. D. Tougaw, W. Porod, and G. H. Bernstein, *Nanotechnology* **4**, 49 (1993).

⁶C. S. Lent and P. D. Tougaw, (unpublished).

⁷Others have attempted to apply our QCA architecture ideas (Refs. 4-6) to a system where the information is encoded in the spin of the electron. See S. Bandyopadhyay, B. Das, and A. Miller (unpublished).

⁸The polarization so defined is not to be confused with a dipole moment. For the situations we consider here, the ground state of the cell has no dipole moment, though it may have a quadrupole moment. The cell polarization simply measures the degree to which the electronic charge is aligned and the direction of that alignment.

⁹For the standard cell, we take E_Q to be the Coulomb energy of two electrons separated by one-third the dot diameter D , a physically reasonable first approximation.

¹⁰Reference 6 examines the behavior of lines of cells as all these Hamiltonian parameters are varied. The bistable behavior we examine here is not critically dependent on a particular choice of these parameters, but holds for a wide range of parameter choices.

¹¹For a review of the current state, see G. J. Ashwell, Ed., *Molecular Electronics* (Wiley, New York, 1992).

¹²Control of cell occupancy is clearly important in implementation. The ability to control quantum dot occupancies over as many as 10^6 dots using a backgating technique has recently been reported by B. Meurer, D. Heitmann, and K. Ploog, *Phys. Rev. Lett.* **68**, 1371 (1992).

¹³P. Bakshi, D. A. Broido, and K. Kempa, *J. Appl. Phys.* **70**, 5150 (1991).

¹⁴See the discussion and two-electron calculation by C. S. Lent, in *Nanostructures and Mesoscopic Systems*, edited by W. P. Kirk and M. A. Reed (Academic, Boston, 1992), p. 183.

¹⁵See the discussion of the behavior of lines of cells as a binary wire in Ref. 6.

¹⁶Enlarging the quantum dots lowers the on-site energy E_0 and makes it energetically more favorable to occupy the larger dots.

¹⁷Noninvasive probing of single-electron charging in a semiconductor quantum dot has recently been reported by M. Field, C. G. Smith, M. Pepper, J. E. F. Frost, G. A. C. Jones, and D. G. Haako, *Phys. Rev. Lett.* **70**, 1311 (1993).

Resonance-antiresonance line shape for transmission in quantum waveguides with resonantly coupled cavities

Wolfgang Porod, Zhi-an Shao, and Craig S. Lent

Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

(Received 24 June 1993)

We investigate the line shape of the transmission probability in quantum waveguides with resonantly coupled cavities. Resonance and antiresonance features in the vicinity of each quasibound state can be characterized by a zero-pole pair in the complex-energy plane, which leads to asymmetrical transmission peaks. We have found a generalization of the familiar symmetrical Lorentzian line shape. Using several examples, we demonstrate the utility of our proposed line shape to extract the lifetime of the quasibound state by a fit to the data. We also discuss the asymmetrical line shapes in the context of Fano resonances.

A problem, which many branches of physics have in common, is to extract lifetimes of quasibound states from peaks in transmission coefficients or scattering cross sections.¹ The underlying theory holds that each quasibound state of the system leads to a pole of the propagator in the complex-energy plane which, in turn, gives rise to a pole in the transmission amplitude. If this pole is sufficiently close to the real-energy axis, it will result in a resonance maximum of the observed transmission coefficient. Particularly well understood is the problem of double-barrier resonant tunneling, where the lifetimes of the quantum-well states may be extracted from the width of Lorentzian-shaped transmission peaks. Less understood is the problem of electronic transport in quantum waveguides with resonantly coupled cavities.² It is known for these structures that the resonator states lead to resonance and antiresonance features,^{3,4} but their detailed line shape has not been investigated so far. In this paper, we present a theory of the line shape for transmission in resonantly coupled quantum waveguides, and we provide a computational method to extract the lifetimes of the corresponding quasibound states.

For double-barrier resonant tunneling (DBRT), it is well known that the resonant transmission phenomena are related to the quasibound states in the quantum-well region.^{5,6} According to the Breit-Wigner theory, the transmission amplitude in the complex-energy plane possesses a pole for each quasibound state.⁷ Specifically, a quasibound state at energy E_p and decay time $\tau = \hbar/(2\Gamma)$ yields a simple pole in the transmission amplitude $t(z)$ at the complex energy $z = E_p - i\Gamma$,

$$t(z) \sim 1/[z - (E_p - i\Gamma)]. \quad (1)$$

If this pole is sufficiently close to the real-energy axis such that the effect of other poles can be neglected, the transmission probability $T(E) = |t(E)|^2$ for a physical energy on the real-energy axis E will be given by

$$T(E) = \Gamma^2 / [(E - E_p)^2 + \Gamma^2]. \quad (2)$$

Here, we have normalized $T(E_p) = 1$ which corresponds to unity transmission in a symmetrical double-barrier structure.⁸ Equation (2) describes a transmission reso-

nance with a Lorentzian line shape. It is an easy matter to extract the lifetime of the quantum-well states from the width of the observed transmission peak.

An example of the transmission amplitude for DBRT is presented in Figs. 1(a) and 1(c), where the double-barrier structure and the transmission channel are schematically displayed in the inset of Fig. 1(a). In this example, the symmetrical barriers are assumed to have a height of $V_0 = 0.2$ eV, thickness of $d = 3$ nm, and separation of $L = 20$ nm. The poles in the complex-energy plane are clearly visible in the contour plot of the absolute value of $t(z)$ shown in Fig. 1(c). The corresponding Lorentzian transmission peaks of the transmission probability $T(E)$ are plotted in Fig. 1(a).

In a recent study,⁹ we have pointed out that the structure of the transmission amplitude in the complex-energy plane is different for resonant quantum waveguide sys-

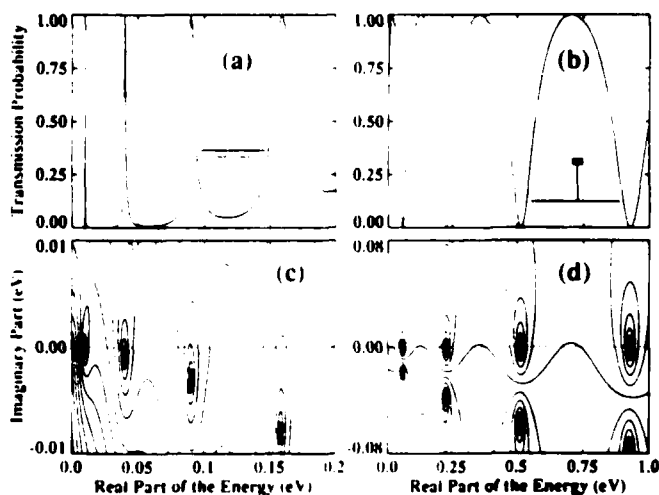


FIG. 1. Comparison of the structure of the transmission amplitude in the complex-energy plane for double-barrier resonant tunneling (poles) and t -stubs (zero-pole pairs). For DBRT, (a) shows the transmission probability on the real-energy axis, and (c) gives a contour plot of the absolute value of the transmission amplitude in the complex-energy plane. For the t -stub structure, the corresponding plots are shown in (b) and (d), respectively.

as compared to the above case of double-barrier resonant tunneling. In essence, each resonator state in a quantum-wire structure gives rise to a zero-pole pair in the complex-energy plane, where the zero *always* occurs on the real-energy axis. As a consequence, there exist transmission zeros, and the zero-pole pairs lead to asymmetric resonance and antiresonance features in the transmission probability with non-Lorentzian line shapes. As an example, we show the behavior for a *t*-stub structure,¹⁰ as schematically shown in the inset of Fig. 1(b), which consists of a main transmission channel and a dangling wire of length $L = 10$ nm. For this so-called strongly coupled *t*-stub,⁹ the transmission probability is shown in Fig. 1(b) and a contour plot of the absolute value of t in the complex-energy plane is displayed in Fig. 1(d). Zero-pole pairs are clearly visible which correspond to the quasibound states in the stub.

Based on the zero-pole pair nature of the resonances, we make the following ansatz for the transmission amplitude in the vicinity of each quasibound state:

$$t(z) \sim (z - E_0) / [z - (E_p - i\Gamma)] . \quad (3)$$

Here, E_0 and $E_p - i\Gamma$ are the positions of the transmission zero and the pole, respectively. The lifetime of the quasibound state is given by $\tau = \hbar / (2\Gamma)$, as for the case of double-barrier resonant tunneling. Again, the transmission probability on the real-energy axis is given by $T(E) = |t(E)|^2$, and the proportionality constant in Eq. (3) is determined by assuming peaks with unity transmission which are known to occur in symmetrical waveguide systems.⁹ (This is no limitation, since the transmission probability always can be rescaled to have a maximum value of one.) A unity transmission peak at energy E_1 provides two constraints for $T(E)$, namely, $T(E_1) = 1$ and $(d/dE)T|_{E_1} = 0$. It is an easy matter to show that

$$T(E) = \Gamma^2 / [(E_p - E_0)^2 + \Gamma^2] \times [(E - E_0)^2 / (E - E_p)^2 + \Gamma^2] . \quad (4)$$

The above expression gives the line shape of the transmission probability for resonantly coupled quantum waveguides in terms of the three parameters, namely, the energy of the transmission zero, E_0 , the energy of the resonant state, E_p (the real part of the pole energy), and the inverse lifetime of the state, Γ (the imaginary part of the pole energy). Note that (4) produces an asymmetrical line with a resonance and antiresonance behavior.

Such asymmetrical line shapes have previously been noted in atomic and molecular physics.¹¹ These so-called Fano resonances are known to occur when a bound state is coupled to a continuum of states, thereby leading to resonance phenomena.¹²⁻¹⁴ In his original paper,¹¹ Fano, after somewhat lengthy derivations, found that the autoionization cross section could be parametrized by $(q + \epsilon)^2 / (1 + \epsilon^2)$, where ϵ is a reduced energy [it is defined as $\epsilon = (E - E_{\text{res}}) / \Gamma$, where E_{res} is the energy of the resonant state] and q is treated as a parameter (it is a complicated expression involving matrix elements). We note that this is the same line shape as our Eq. (4) by making the following substitutions: $\epsilon = (E - E_p) / \Gamma$ and

$q = (E_p - E_0) / \Gamma$. Comparing our approach to Fano's,¹¹ we note that ϵ has a similar meaning where E_p , the real part of the pole energy, corresponds to E_{res} , the energy of the resonant state. For the parameter q , our approach yields a simple expression which could not have been inferred from Ref. 11. Apparently, Fano's line shape corresponds to a zero-pole structure in the complex-energy plane, a fact which has not been noted before.

Given a certain transmission curve, we now can fit each resonance and antiresonance feature to obtain the lifetime of the corresponding quasibound state. Using the known energies of the transmission zero, E_0 , and transmission one, E_1 , we can determine the energy of the pole as

$$E_p = (E_0 + E_1) / 2 \pm \sqrt{(E_0 - E_1)^2 - 4\Gamma^2} / 2 . \quad (5)$$

The choice of the sign in the above equation determines whether $E_p > E_0$ or $E_p < E_0$. With this, the only unknown parameter is Γ which may be used to obtain the best fit of the theoretical line shape (4) to the given transmission curve. We seek the best fit in the sense of the least mean square error.

We now present several examples of fits to the line shapes of resonance and antiresonance pairs, each corresponding to a quasibound state of a quantum waveguide structure. We also extract the lifetime of the state. In essence, we extract information about the imaginary part of the pole from data on the real-energy axis. Since in our examples we know the transmission amplitude in the entire complex plane, we can ascertain the quality of the fit.

Figure 2 presents fits of the resonance and antiresonance line shapes for a family of so-called weakly coupled *t*-stubs,⁹ which are schematically shown in the insets. In Figs. 2(a), 2(b), and 2(c), the resonant stubs are separated from the main transmission channel by a tunneling barrier of length $l = 1.0$ nm and height $V_0 = 0.5, 1.0$, and 2.0 eV, respectively. In each case, we show three quasibound states which lead to zero-one features in the transmission probability, and which are labeled in the plots. Figures 2(a), 2(b), and 2(c) show the fitted line shape for the resonance numbered three of cases 2(a), 2(b), and 2(c), respectively. The fit is shown by the dotted line, and the curve to be fitted by the solid line. Note the extremely good fit of our ansatz, Eq. (3). In Table I, we give a comparison of the positions of the fitted poles and the exact poles for the weakly coupled *t*-stubs shown in Fig. 2. For a particular pole, the difference between the fitted and the exact values is due to the interaction with other poles.

Figure 3 presents another example which shows transmission for a *t*-stub in addition to double-barrier resonant tunneling on the main transmission channel. A schematic drawing of this waveguide structure is displayed in the inset, and the two tunneling barriers have a thickness of 1 nm, height of 0.5 eV, and separation of 4 nm. Figure 3(a) shows the transmission probability on the real-energy axis, and we see various types of resonance behavior. Resonances 1 and 4 exhibit more or less sharp $1 \rightarrow 0$ transition, whereas resonances 2 and 3 are of the $0 \rightarrow 1$ type. Figures 3(b) – 3(e) display the fits according to our ansatz, Eq. (3), for resonances 1–4, respective-

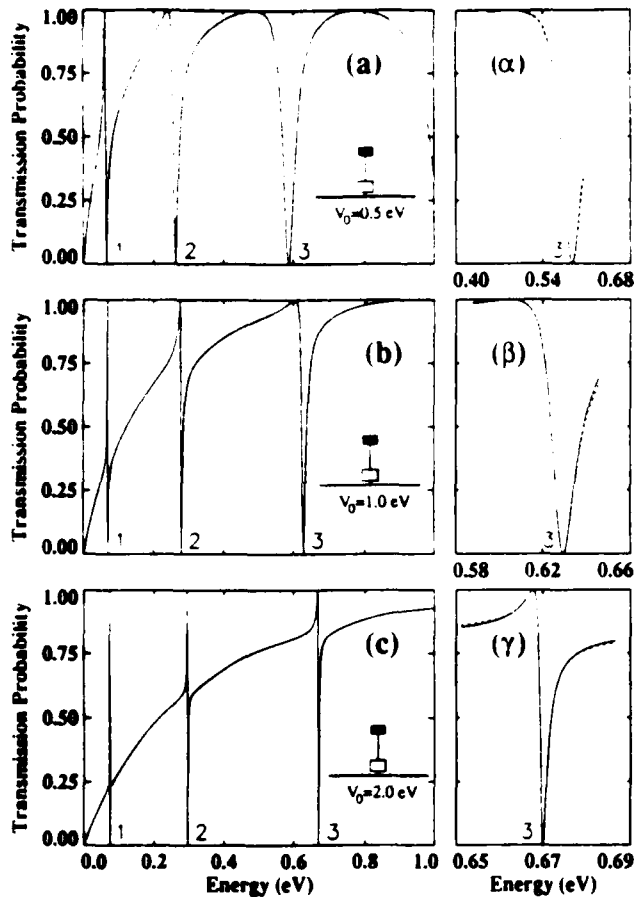


FIG. 2. Examples of fits using the line shape (4) for the weakly coupled t -stub structures shown in the insets. The fits to the third resonant state of (a), (b), and (c) are given in (α) , (β) , and (γ) , respectively. The fits are shown by the dotted line, and the curve to be fitted by the solid line.

ly. Again, the fit is shown by the dotted line, and the curve to be fitted by the solid line. It appears that the zero-pole character of each quasibound state leads to extremely good fits of the transmission probability in the vicinity of each resonance.

In general, the locations of the poles and the zeros on the real-energy axis are not the same, i.e., $E_p \neq E_0$. It is this fact that gives rise to the asymmetric line shape, Eq. (4). Note that from Eq. (5), the position of the pole E_p is always between the positions of the transmission zero and

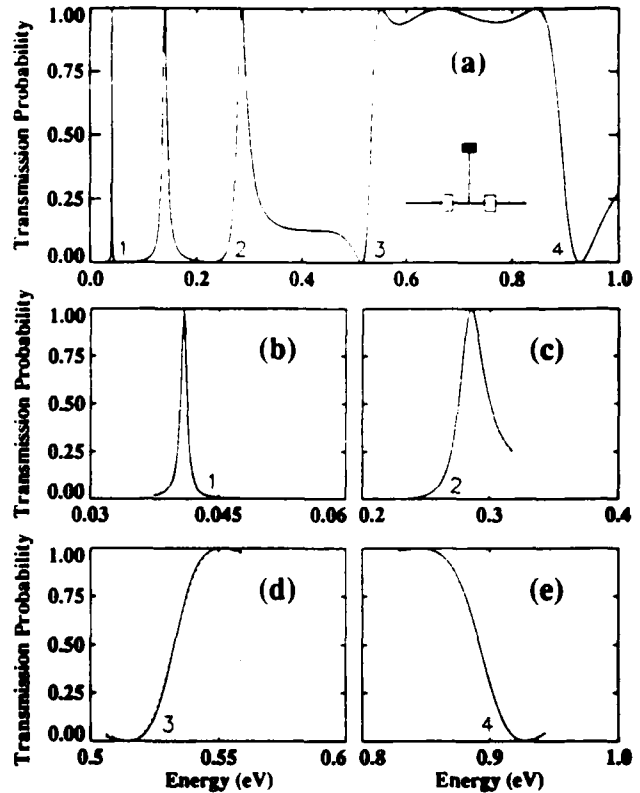


FIG. 3. Example of fits using the line shape (4) for a t -stub with a double-barrier structure on the transmission channel shown in the inset to (a). The fits of resonances 1–4 are shown in (b)–(e), respectively.

one, E_0 and E_1 . For $E_p > E_0$, the transmission resonance is of the $0 \rightarrow 1$ type. A $1 \rightarrow 0$ transition results for $E_p < E_0$. If the pole and the zero occur at the same real energy, i.e., $E_p = E_0$, then Eq. (4) yields a symmetric line shape,

$$T|_{E_p=E_0} = (E - E_p)^2 / [(E - E_p)^2 + \Gamma^2] \quad (6)$$

$$= 1 - \Gamma^2 / [(E - E_p)^2 + \Gamma^2]. \quad (7)$$

The above expression describes a Lorentzian-shaped reflection line.

In recent work,¹⁵ Price has pointed out that a resonant quasibound state can give rise to either Lorentzian-shaped transmission or reflection peaks, and the terms

TABLE I. The exact and fitted positions of the transmission poles for the weakly coupled t -stubs shown in Fig. 2.

Structure	Zero index	Exact pole (eV)	Fitted pole (eV)
$V_0 = 0.5$ eV	1	0.064 305 – 0.001 686 <i>i</i>	0.064 321 – 0.001 659 <i>i</i>
	2	0.260 208 – 0.007 807 <i>i</i>	0.259 942 – 0.007 715 <i>i</i>
	3	0.585 195 – 0.020 514 <i>i</i>	0.583 005 – 0.021 300 <i>i</i>
$V_0 = 1.0$ eV	1	0.069 532 – 0.000 329 <i>i</i>	0.069 529 – 0.000 327 <i>i</i>
	2	0.278 960 – 0.001 985 <i>i</i>	0.278 965 – 0.001 971 <i>i</i>
	3	0.627 876 – 0.006 535 <i>i</i>	0.627 752 – 0.006 525 <i>i</i>
$V_0 = 2.0$ eV	1	0.073 625 – 0.000 031 <i>i</i>	0.073 625 – 0.000 031 <i>i</i>
	2	0.295 804 – 0.000 227 <i>i</i>	0.295 801 – 0.000 228 <i>i</i>
	3	0.669 735 – 0.000 857 <i>i</i>	0.669 739 – 0.000 852 <i>i</i>

these peaks resonances of the first and the second kind, respectively. We see that the reflection peaks in general will not have a Lorentzian shape, and that Price's resonances of the second kind are recovered when $E_p = E_0$.

In addition to fitting the line shape for experimental data, we anticipate that our expression (4) will also be useful in extracting the lifetimes of resonant states for calculated curves. We have in mind circumstances where one only calculates the transmission probability on the real-energy axis because it is computationally too expensive to determine the transmission amplitude in the complex-energy plane.^{3,14}

In summary, we have investigated the detailed line shape of the transmission probability in quantum

waveguides with resonantly coupled cavities. The resonance and antiresonance features in the vicinity of each quasibound state can be characterized by a zero-pole pair in the complex-energy plane. We have found a generalization of the familiar symmetrical Lorentzian resonance peaks. Using several examples, we have demonstrated the utility of our asymmetrical line shape (4) to extract the lifetime of the quasibound state by a fit to the data. We also discussed the asymmetrical line shapes in the context of Fano resonances.

This work was supported, in part, by AFOSR, ARPA, and ONR.

¹L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon, New York, 1962).

²*Nanostructure Physics and Fabrication*, edited by M. A. Reed and W. P. Kirk (Academic, Boston, 1989).

³C. S. Lent, in *Computational Electronics: Semiconductor Transport and Device Simulation*, edited by K. Hess, J. P. Leburton, and U. Ravaioli (Kluwer, Boston, 1991), p. 259.

⁴P. J. Price, IEEE Trans. Electron. Devices 39, 520 (1992).

⁵P. J. Price, Phys. Rev. B 38, 1994 (1988).

⁶M. Büttiker, IBM J. Res. Dev. 32, 63 (1988).

⁷T. B. Bahder, C. A. Morrison, and J. D. Bruno, Appl. Phys. Lett. 51, 1089 (1987).

⁸Gaston García-Calderón, Solid State Commun. 62, 441 (1987).

⁹W. Porod, Z. Shao, and C. S. Lent, Appl. Phys. Lett. 61, 1350

(1992).

¹⁰F. Sols, M. Macucci, U. Ravaioli, and K. Hess, Appl. Phys. Lett. 54, 350 (1989); J. Appl. Phys. 66, 3892 (1989); S. Datta, Superlatt. Microstruct. 6, 83 (1989); S. Subramaniam, S. Bandyopadhyay, and W. Porod, J. Appl. Phys. 68, 4861 (1990); Y. Takagaki and D. K. Ferry, *ibid.* 72, 5001 (1992); J.-B. Xia, Phys. Rev. B 45, 3593 (1992).

¹¹U. Fano, Phys. Rev. 124, 1866 (1961).

¹²E. Tekman and P. F. Bagwell, Phys. Rev. B 48, 2553 (1993).

¹³T. B. Boykin, B. Pezeshki, and J. S. Harris, Jr., Phys. Rev. B 42, 12 769 (1992); T. B. Boykin, *ibid.* 47, 12 696 (1993).

¹⁴D. Z.-Y. Ting and T. C. McGill, Phys. Rev. B 47, 7281 (1993).

¹⁵P. J. Price, Appl. Phys. Lett. 62, 289 (1993).

Spatial density of lines exposed in poly(methylmethacrylate) by electron beam lithography

Xiaokang Huang, Gary H. Bernstein, and Greg Bazan

Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

Davide A. Hill

Department of Chemical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

(Received 30 September 1992; accepted 8 February 1993)

In this article we report the experimental investigation of the density of patterns exposed with electron beam lithography (EBL). A linear relationship was found between the minimum width of poly(methylmethacrylate) (PMMA) walls and the given resist thickness (height of the walls). Below the minimum width, PMMA walls become wavy or fail due to the internal stresses caused by PMMA swelling during development. Previous research pertaining to proximity effects at feature sizes and spacing below the $0.25\text{ }\mu\text{m}$ range is limited. We have experimentally investigated proximity effects in very-high-density gratings in the pitch range from 50 to 330 nm. The relationship between electron beam dose and grating pitch, where proximity effects during electron beam lithography play a major role, was achieved. By fitting the experimental results with a triple Gaussian model, the contributions of the different electron distributions in proximity effects were determined. It was found that fast secondary electrons dominate the proximity effects in the range we studied and they limit the density of patterns fabricated by EBL.

I. INTRODUCTION

The fabrication of nanostructures, and in particular quantum devices, depends on the resolution and achievable pattern density of lithographic techniques. In electron beam lithography (EBL), the lithographic resolution relies on that of the electron beam generator and the contrast of the resist/developer system. Poly(methylmethacrylate) (PMMA) is still the most popular choice of resist for nanostructure fabrication because of its extremely high resolution.

In the lift-off process commonly used in nanofabrication, PMMA exposed with an electron beam is followed by a develop step, after which the unexposed resist remains on the substrate and acts as a shadow for metal evaporation. For very-high-density patterns, the remaining PMMA is in the form of a thin wall between the developed lines, which serves as the spacer between deposited metal lines. Although PMMA is usually considered to be a nonswelling resist, accounting partially for its high-resolution properties, some absorption of developer is necessary to aid in the development process.¹ A very small amount of swelling can induce internal stresses in thin PMMA walls which weaken its mechanical properties and cause the deformation of thin resist walls during development in the fabrication of very close lines. In the limit of very-high-resolution lithography, lines are placed so closely together that the resist walls can either become wavy, fall over, or fail completely.^{2,3} Therefore, the buckling of the PMMA walls limits the ultimate density of patterns in the EBL process. Although Chen² predicted that the limit of line pitch is $\sim 75\text{ nm}$, our results show that for 100-nm-thick PMMA walls, PMMA can be as narrow as 20 nm without resist failure,³ resulting in higher density than expected by them.

Sub-50-nm pitch semiconductor-metal photodetectors have also been achieved by Chou *et al.*⁴

Another serious limitation on very-high-density patterns with electron beam lithography is the proximity effect. This is due to the distribution of the primary electron beam, forward scattered electrons, secondary electrons scattered in the resist film on the substrate, and backscattered electrons from the substrate. It is important because those scattered electrons expose regions of the resist that are not originally written by the electron beam. For the schematic grating shown in Fig. 1(a), the electron beam exposes the pattern on each line. In the ideal case, the resulting energy distribution in the resist is very spatially confined [Fig. 1(b)]. However, the actual energy distribution for very-high-density gratings is less distinct,⁵ as shown in Fig. 1(c), and proximity effects decrease the modulation transfer function of the exposure for closely spaced patterns. [The modulation transfer function (MTF) is defined in Fig. 1(c).] Assume the peak dose received by each line is E_p and the lowest dose received at the areas between lines is E_v . The resolution of gratings depends on the capability of a given developer to distinguish between E_v and E_p . The distribution of electrons in *e*-beam lithography is complicated by the fact that it depends on almost all parameters in lithography, e.g., accelerating potential, beam size, photoresist thickness and type, and substrate properties. Investigations of the energy distributions have been reported elsewhere.⁵⁻¹⁰ A double Gaussian model has mostly been used for the electron distributions.^{5,6,11-14} Computer-aided proximity effect correction and other methods reducing the proximity effects during *e*-beam lithography have been pursued.¹⁵⁻²⁰ The feature size and spacing in most of the published work are in the half-micron regime, which is quite useful for fabri-

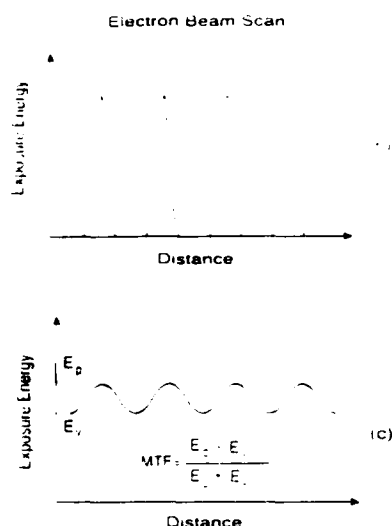


FIG. 1. Exposure energy distribution in very-high-density gratings. (a) Schematic of the beam scan for a grating. (b) Energy distribution in the near ideal case for the beam scan shown in (a). (c) Actual energy distribution in very-high-density gratings.

cating ULSI circuits. In this case, backscattered electrons dominate the proximity effects because of the typically low beam energies utilized (~ 25 keV). The width of the backscattered electron distribution is in the order of the feature size and spacing. However, previous research pertaining to proximity effects at feature sizes and spacing below the $0.25\text{ }\mu\text{m}$ range is limited.⁹⁻¹¹

In this paper, we report the experimental investigations of the density of the electron beam lithography limited by the physical strength of PMMA and proximity effects. Proximity effects in very-high-density gratings were experimentally studied in the pitch range from 50 to 330 nm. We will discuss the relationship between electron beam dose and grating pitch, where proximity effects during electron beam lithography play a major role.

II. EXPERIMENTAL PROCEDURE

The EBL system used in this study consisted of an Amray 1400 scanning electron microscope (SEM) with a maximum beam energy of 50 keV and a minimum beam width of 5 nm using a W cathode, controlled by an IBM PS/2 personal computer interfaced through a Pragmatic Instruments 2201A 16-bit arbitrary waveform generator.²¹ Special care was taken to minimize all noise sources. PMMA (950 000 amu) was spun on Si and SiO_2/Si wafers and baked at 170°C for 4 h. Mixtures of methylisobutylketone (MIBK):2-propanol (IPA) (1:3) with the addition of 1% and 1.5% methylethylketone (MEK) by volume, were used to develop samples. This mixture has been shown to have very high development contrast properties.²² For both the experiments on proxim-

TABLE I. Experimental parameters.

Experiment Patterns	Proximity effect Arrays of gratings	Strength of PMMA Two parallel lines or gratings
Line pitch	50–330 nm	70–150 nm
Accelerating voltage	50 kV	45 kV
Resist thickness	60–70 nm	105–180 nm
Developer	MIBK:IPA:MEK = 1:3:0.06 (1.5% MEK)	MIBK:IPA:MEK = 1:3:0.04 (1% MEK)
Developer temp. ($^\circ\text{C}$)	26	25
Developing time	6–10 s	15 s
Metal thickness	15 nm	30 nm

ity effect and the strength of PMMA, different process parameters were used. A summary of the experimental parameters is given in Table I. The resist thickness in the proximity effect measurement was kept to 60–70 nm in order to avoid PMMA failure during the processes. Gratings with different pitch in the proximity effect experiments are $3\text{ }\mu\text{m} \times 3\text{ }\mu\text{m}$ in size and separated by $2\text{ }\mu\text{m}$ with each other in order to eliminate the proximity effects caused by the scattered electrons from adjacent gratings.²³

III. RESULTS AND DISCUSSION

A. Strength of PMMA

In order to achieve liftoff of evaporated metal films, it is necessary that PMMA walls remain between the developed areas. The width of PMMA walls was determined by measuring the spaces between resulting metal patterns after liftoff. The undercutting for thin resists was neglected because it is small for thin resists and high beam energies. The height of the walls after development is assumed to be the same as the initial resist thickness because the unexposed PMMA of very high molecular weight has very low solubility.²² One might suppose that in the case of high-density gratings, areas between lines might be partially developed because of proximity effects. For grating pitch larger than 70 nm, we find in the next section that $E_v/E_p < 0.6$. For the developer with contrast, γ^* , larger than 10, the amount of the resist that is removed through developing is small compared with the original thickness of the resist, and therefore can be ignored in the final resist thickness determination.

PMMA can be considered a strong, glassy material after baking. However, during development, the absorption of the developer causes swelling, and internal stresses result in the deformation of PMMA walls. The wall aspect ratio (height-to-width) is an important parameter in delineating the stability threshold, since a tall, thin wall will be strongly susceptible to small dimensional perturbations around the erect position, while a short and thick wall will be able to sustain a higher degree of swelling before buckling.

Figure 2 shows experimental results of the minimum width of a PMMA wall achievable for a given starting resist thickness. Error bars represent variations in metal line separation (and therefore wall thickness) over the length of the lines. These data were obtained by smoothly

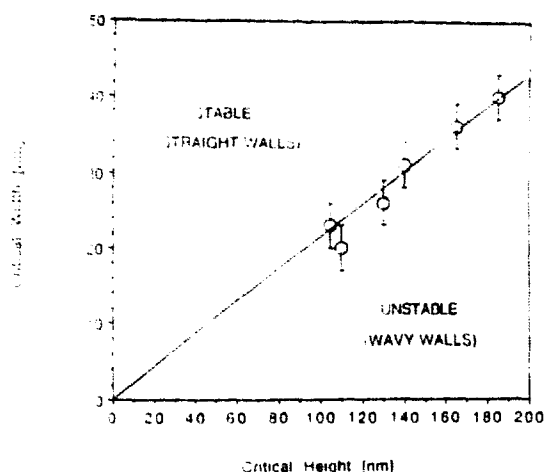


FIG. 2. Minimum width of PMMA walls achievable vs starting resist thickness (wall height).

varying the line pitch for a given resist thickness and observing the point at which lines either failed completely or became noticeably wavy. The data indicate a direct proportionality between critical wall height and width, implying a constant aspect ratio for buckling. The area above the critical line is the regime in which straight PMMA walls can be achieved. The area below it is the regime in which the walls become wavy or fail.

The stability of an infinitely long wall of width w and height h with a rigidly anchored bottom and a top free end can be modeled with linear plate theory.²⁴ The details of the theoretical calculation are given in Ref. 25. According to this model, the critical wall aspect ratio for stability, $(h/w)^*$, is given as a function of the swelling strain:

$$\left(\frac{h}{w}\right)^* = \frac{13}{\sqrt{15(1-\nu^2)\epsilon_{sw}}} \quad (1)$$

where ν is Poisson's modulus and ϵ_{sw} is the hypothetical swelling strain that the wall would undergo if allowed to expand freely. For $(h/w) > (h/w)^*$ the wall will buckle as shown in Fig. 3(a). The bright areas in Fig. 3 are evaporated gold, and the dark areas are the areas where PMMA walls stood before liftoff. Equation (1) predicts a linear relationship between the critical wall height and width for a constant ϵ_{sw} , which is consistent with our experimental results as shown in Fig. 2. From the slope of the line in Fig. 2, the critical aspect ratio for stability can be calculated, giving $(h/w)^* = 4.75$. Then $\epsilon_{sw} = 5.1\%$ for $\nu = 0.5$ from Eq. (1). This indicates that 5% swelling can lead to failure of PMMA walls with aspect ratio greater than 5.

In making gratings, a set of parallel PMMA walls forms during development. The buckling of one of the walls can lead to the catastrophic failure of the whole grating, because surface forces can induce cross correlation between deformation of adjacent structures, producing a pattern of adjacent sinusoidal waves 180° out of phase with respect to each other as shown in Fig. 3(b). Figure 3(c) shows a 50-nm pitch grating with straight resist walls between lines.

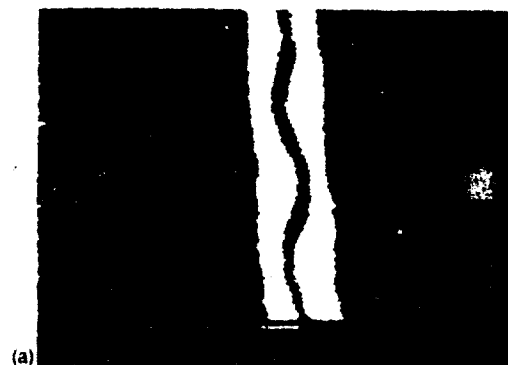


FIG. 3. Remaining metal patterns after liftoff. (a) Failure of parallel lines caused by wavy PMMA walls between them. (b) A typical failure of gratings resulting from PMMA wall swelling and buckling. (c) A 50 nm grating with straight PMMA walls between lines. Light areas are gold, and dark areas are the shadows of PMMA walls which have been removed during liftoff.

B. Proximity effects

Proximity effects must be strongly considered in the fabrication of very-high-density patterns. It can be expected that when lines get closer together, the dose needed for exposure should decrease because interline scattering of electrons causes additional energy deposition. Experimental results of normalized dose versus grating pitch are shown in Fig. 4 along with theoretical simulations. In the

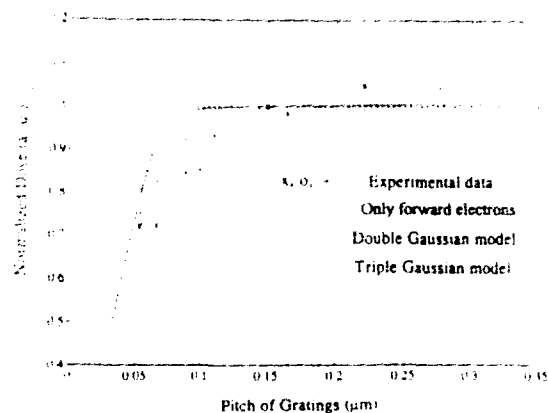


FIG. 4. The relationship between normalized dose and grating pitch. Doses are normalized to the dose required for the same linewidth at a pitch of 150 nm.

figure, doses for various pitch values are normalized to that needed for the same linewidth in the 150 nm pitch grating. Three samples with different doses and development time were measured. The data of all three samples are very consistent. Besides exposure dose, many factors affect linewidth in experiments, such as contrast of developers, developing time, and temperature. Our results show that although linewidth may change with other factors, their effects on the relationship between normalized dose and grating pitch are small and can be neglected. It can also be seen that normalized dose decreases slowly with pitch until about 100 nm when the change becomes rapid. As we will show, this effect is due to the different distributions of scattered electrons in the resist during exposure. Models of the exposure energy distribution in the resist employed for fitting the experimental data include the double Gaussian⁶:

$$f(r) = \frac{1}{\pi(1+\eta)} \left[\frac{1}{\alpha^2} \exp\left(-\frac{r^2}{\alpha^2}\right) + \frac{\eta}{\beta^2} \exp\left(-\frac{r^2}{\beta^2}\right) \right], \quad (2)$$

and the triple Gaussian model⁷:

$$f(r) = \frac{1}{\pi(1+\eta+\eta')} \left[\frac{1}{\alpha^2} \exp\left(-\frac{r^2}{\alpha^2}\right) + \frac{\eta}{\beta^2} \exp\left(-\frac{r^2}{\beta^2}\right) + \frac{\eta'}{\gamma^2} \exp\left(-\frac{r^2}{\gamma^2}\right) \right]. \quad (3)$$

The simulation curves are shown in Fig. 4. In Eqs. (2) and (3) the first term is used to describe the primary beam and forward scattered electron distribution. The second term accounts for the behavior of the backscattered electrons. The third term in the triple Gaussian model is added to describe all other exposures that are not included in the first two terms. The third term may result from large angle forward scattered electrons, secondary electrons, or even broad tails in the primary beam distribution.^{8,9} The results of Monte Carlo calculations by Joy¹ and Murata *et al.*² clearly show broad tails to the energy distribution due to fast secondary electrons, which we will show play an important role in proximity effect calculations over the spatial

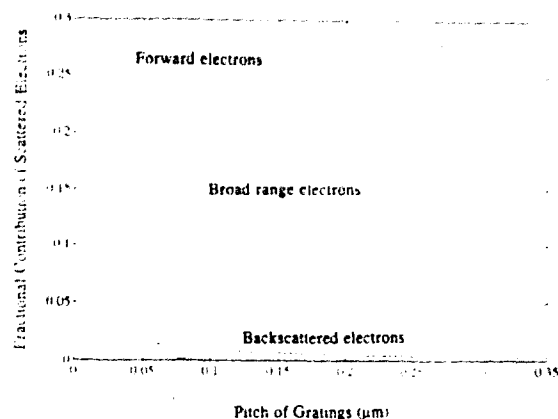


FIG. 5. Interline proximity effects caused by scattered electrons vs pitch of gratings.

range discussed here. In Eqs. (2) and (3), α , β , and γ are the widths of each Gaussian distribution, and η and η' are the ratios of exposures of the second and third term to the forward exposure, respectively. According to Ref. 8, $\beta = 10 \mu\text{m}$ and $\eta = 0.8$ for 50 keV electrons in solid silicon substrates, α , γ , and η' were chosen for the apparent best fit. The dashed line in Fig. 4 shows the calculation in which only the first Gaussian term, that is, only the contribution of forward scattered electrons, was considered. Since the width of this term is only a few tens of nanometers, forward scattered electrons only cause a short range proximity effect.

When the second Gaussian term was added in the calculation (the double Gaussian model), the weak change of dose over large changes in pitch (solid line in Fig. 4) was caused by the long-range proximity effect of backscattered electrons. The small difference between the dashed curve and the solid curve shows that the proximity effect caused by backscattered electrons at 50 keV is weak. $\alpha = 0.037 \mu\text{m}$ was used in these two curves. The dash-dot line in Fig. 4 was calculated with the triple Gaussian model. This curve fits the data very well with $\alpha = 0.029 \mu\text{m}$, $\gamma = 0.4 \mu\text{m}$, and $\eta' = 0.35$. We conclude, therefore, that the proximity effect from the γ -term electrons is very important in high-density patterns in EBL. Since these electrons cause proximity effects in a broad range relative to the forward scattered electrons, and their sources are still not clear, we call them simply "broad range electrons."

The ratios of the contributions of scattered electrons from the interline exposures to the total exposure received by each line are shown in Fig. 5. The solid line, the dashed line, and dash-dot line correspond to the interline exposure contributions of forward scattered electrons, backscattered electrons, and broad range electrons, respectively. From Fig. 5 we see that for the interline proximity effects: (i) The contributions of backscattered electrons weakly depends on the grating pitch because of their wide distribution [$\sim 10 \mu\text{m}$ (Refs. 8 and 26)] at 50 keV. The backscattered coefficient is a constant for certain materials when the beam energy is larger than 20 keV.²⁷ Hence the intensity of

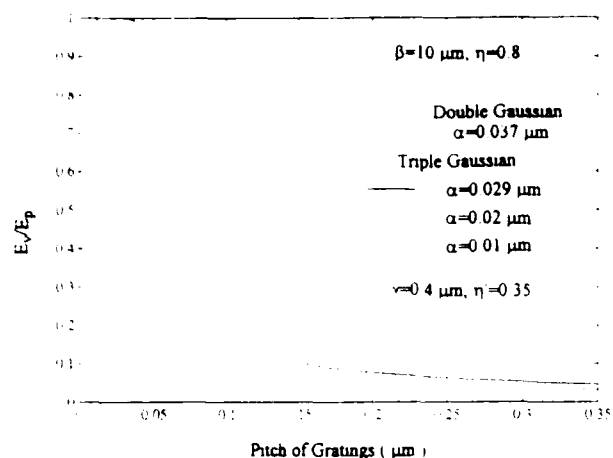


FIG. 6. E/E_p vs grating pitch. E is the lowest exposure energy received in the areas between lines. E_p is the maximum exposure energy received by each line

the backscattered electron distribution for 50 keV is much lower than for 20 keV because it spreads much further. Therefore, for high-density patterns, the proximity effect of backscattered electrons is like a weak background. It can be ignored in the total exposure because the contribution of backscattered electrons is less than 2% and much smaller than the contribution of the broad range electrons in the whole pitch range over which our calculations were performed. (ii) The effect of forward scattered electrons is not important until the pitch of gratings approaches the width of its distribution. Their effect increases dramatically when the pitch is smaller than 50 nm. Therefore, the proximity effect of forward electrons will limit the resolution of electron beam lithography. (iii) Broad range electrons dominate the proximity effect in the pitch range from 50 to 350 nm or even larger in our experiments. This range is on the order of feature sizes of quantum devices and ULSI circuits. Therefore it is very important to include this term in the proximity effect correction. Since the contribution of the broad range electrons to the total exposure is larger than 20% when grating pitch is less than 50 nm, their limitation on the resolution and density of electron beam lithographically defined patterns cannot be ignored.

We further consider here the role of fast secondary electrons in proximity effects. Fast secondary electrons have a complicated distribution; it is a narrow peak with a wide shoulder.^{5,7} The width of the narrow peak is about 10 nm in free-standing resist. Its role in proximity effects was included in the forward scattered electrons⁵ which are described by the first term of the double or triple Gaussian models. However, the proximity effect caused by the wide shoulder was neglected by previous researchers. From Ref. 5 we note that the half-width of the shoulder is about 100 nm and is on the order of the distribution of the broad range electrons. Therefore, it can be concluded that fast secondary electrons are part of the broad range electrons, but they are not the only part. Therefore, the fast secondary electrons introduce both short- and long-range proximity effects, and thereby limit the resolution of EBL.

As mentioned above, the total dose at the region between the electron beam scanned areas increases when lines are placed closer together. The resolution of gratings created with EBL depends on how well a developer can distinguish between the deposited energy densities between lines and at the lines, E and E_p , respectively. Figure 6 shows E/E_p as a function of grating pitch with different Gaussian parameters. The dotted line results from the double Gaussian model with the same parameters used in the calculation of Fig. 4. For 50-nm pitch gratings, the ratio E/E_p is about 0.98. This small difference between exposure energies cannot be distinguished with any practical developers, in contradiction to the experimental results. Therefore, although the double Gaussian model can be used to qualitatively simulate the normalized dose changing with grating pitch, it fails in predicting the density and resolution limitation in electron beam lithography at these very small dimensions. The solid line is calculated with the triple Gaussian model with the same parameters that fit the experimental data in Fig. 4. It can be seen that E/E_p approaches unity when the grating pitch decreases. $E/E_p = 1$ means that the grating patterns are completely washed out because of the proximity effects. For 50-nm pitch gratings, $E/E_p \approx 0.87$. The contrast, γ^* , of our developer (1.5% MEK) is larger than 10 and corresponds to $D_i/D_f \approx 0.8$,²² where D_i is defined as the critical dose below which the resist is not developed at all, and D_f is the dose above which the resist is totally dissolved. (Note that the exposure energy E is proportional to the received dose D so contrast curves yield useful information about the limits on absorbed energy.) The resist will be partially dissolved when the dose is less than D_f , but greater than D_i . This means that the height of the PMMA walls was decreased for 50-nm pitch gratings during development because of partial developing. Since the initial PMMA thickness in our case was 60–70 nm, the height of remaining PMMA walls is assumed to be larger than 40 nm, which we have found is sufficient for good liftoff of 15 nm metal patterns. This also shows that the development is very critical in making very high density patterns; a slight overdevelopment may cause the PMMA walls to be dissolved.

From Fig. 5 we see that forward scattered electrons, including fast secondary electrons, cause the proximity effects to dramatically increase when grating pitch approach α , the width of the distribution of forward scattered electrons. Thus the resolution of EBL can be improved when α decreases. This is clearly shown by the dashed line ($\alpha = 20$ nm) and the dash-dot line ($\alpha = 10$ nm) in Fig. 6. This requires using a very-high-energy system with a non-Gaussian-shaped beam or a resist that depends less on low energy electrons in forward electron scattering. Assuming $\alpha = 10$ nm for a very high-resolution electron beam, a grating with 20 nm pitch can be expected using a developer with the contrast $\gamma^* = 10$, as shown in Fig. 6. This agrees with predictions of Joy⁶ that 10 nm represents the ultimate resolution for PMMA as an electron beam resist. Our results restate those of Joy that the narrow peak of the energy deposited by secondary electrons is practically flat out to a distance of 5 nm, and therefore so low in contrast

(over the peak) as to be indistinguishable to developers with reasonable contrast values.

We saw that the pattern density is also limited by the failure of PMMA walls. As discussed above, the lower the achieved thickness of the PMMA walls, the closer is the achievable line spacing. However, the resist thickness cannot be lowered arbitrarily because, for good liftoff, the resist thickness should be about three times the metal thickness, which depends on the ultimate desired device characteristics. According to Fig. 2, 10 nm spacing between metal lines can be achieved with 40-nm-thick PMMA, and is sufficient for successful liftoff of 10 to 15-nm-thick metal lines. Therefore, we conclude that the limits to the density of metal lines due to the proximity effect and the strength of PMMA walls are consistent for the achievement of lines with pitch as small as 20 nm.

IV. SUMMARY

The proximity effects in electron beam lithography were experimentally investigated with different grating pitch. The results were successfully simulated with a triple Gaussian model. Proximity effects caused by backscattered electrons were insignificant in our case because of their wide distribution. Forward scattered electrons were very important when the grating pitch was about the width of this electron distribution. Broad range electrons, described by the third term in the triple Gaussian model, dominate the proximity effect in the pitch range from 50 to 350 nm. Secondary electrons cause both short- and long-range proximity effects. Their effects can be included in forward scattered electrons as well as broad range electrons. Their limitation on the EBL resolution cannot be ignored.

Upon exposure to suitable solvents, such as common developing and rinsing solutions, narrow PMMA walls were found to become unstable as a result of two effects: (1) swelling, and (2) lateral surface forces between adjacent structures. We investigated the relationship between minimum line spacing and maximum PMMA thickness for EBL fabrication using a high contrast developer. We also showed experimental evidence that supported our predictions, thus demonstrating the existence and severity of such instabilities in nanolithography.

From our calculation of E_r/E_p as a function of grating pitch with the triple Gaussian model, a 20-nm pitch grating can be expected using a very high-resolution electron beam generator and high contrast resist/developer system. The closest spacing between metal lines can be as narrow as 10 nm for 10 to 15-nm-thick metals, implying that 20 nm gratings with 10 nm linewidth and 10 nm line spacing are achievable.

ACKNOWLEDGMENTS

The authors thank D. T. Leighton for helpful discussions, and S. Subramaniam for help in sample preparation and useful discussion. We gratefully acknowledge IBM, AFOSR, ONR, and the University of Notre Dame for their support of this research. D. A. H. thanks the Donors of the Petroleum Research Fund, administered by the American Chemical Society, for partial support of this research.

- ¹A. Weill, in *The Physics and Fabrication of Microstructures and Microdevices*, edited by M. J. Kelly and C. Weisbuch (Springer, New York, 1986), p. 58.
- ²Z. W. Chen, in *Advances in Electronics and Electron Physics*, edited by Peter W. Hawkes (Academic, New York, 1992), Vol. 83, p. 187.
- ³X. Huang, G. Bazan, D. A. Hill, and G. H. Bernstein, *J. Electrochem. Soc.* **139**, 2952 (1992).
- ⁴S. Y. Chou, Y. Liu, and P. B. Fischer, *IEDM* **91**, 745 (1991).
- ⁵D. C. Joy, *Microelectron. Eng.* **1**, 103 (1983).
- ⁶T. H. P. Chang, *J. Vac. Sci. Technol.* **12**, 1271 (1975).
- ⁷K. Murata, D. F. Kyser, and C. H. Ting, *J. Appl. Phys.* **52**, 4396 (1981).
- ⁸R. M. Mankiewich, L. D. Jackel, and R. E. Howard, *J. Vac. Sci. Technol. B* **3**, 174 (1985).
- ⁹S. J. Wind, M. G. Rosenfield, G. Pepper, W. W. Molzen, and P. D. Gerber, *J. Vac. Sci. Technol. B* **7**, 1507 (1989).
- ¹⁰R. J. Bojko and B. J. Hughes, *J. Vac. Sci. Technol. B* **8**, 1909 (1990).
- ¹¹S. A. Rishton and D. P. Kern, *J. Vac. Sci. Technol. B* **5**, 135 (1987).
- ¹²A. S. Chen and A. R. Neureuther, *J. Vac. Sci. Technol. B* **3**, 148 (1985).
- ¹³A. Misaka, K. Harafuji, and N. Nomura, *J. Appl. Phys.* **68**, 6472 (1990).
- ¹⁴G. Owen, *J. Vac. Sci. Technol. B* **8**, 1889 (1990).
- ¹⁵G. Owen and P. Rissman, *J. Appl. Phys.* **54**, 3573 (1983).
- ¹⁶S. W. J. Kuan, C. W. Frank, Y. H. Yen Lee, T. Eimori, D. R. Allee, R. F. W. Pease, and R. Browning, *J. Vac. Sci. Technol. B* **7**, 1745 (1989).
- ¹⁷D. G. L. Chow, J. F. McDonald, D. C. King, W. Smith, K. Molnar, and A. J. Steckl, *J. Vac. Sci. Technol. B* **1**, 1383 (1983).
- ¹⁸R. L. Kostelak, E. H. Kung, M. G. R. Thomson, and S. Vaidya, *J. Vac. Sci. Technol. B* **6**, 2042 (1988).
- ¹⁹E. Nishimura, T. Takigawa, T. Abe, and Y. Katoh, *J. Vac. Sci. Technol. B* **4**, 164 (1986).
- ²⁰R. C. Frye, *J. Vac. Sci. Technol. B* **9**, 3054 (1991).
- ²¹G. Bazan and G. H. Bernstein, *J. Vac. Sci. Technol. A* **11**, 1745 (1993).
- ²²G. H. Bernstein, D. A. Hill, and W.-P. Liu, *J. Appl. Phys.* **71**, 4066 (1992).
- ²³K. L. Lee and H. Ahmed, *J. Vac. Sci. Technol.* **19**, 946 (1981).
- ²⁴S. P. Timoshenko and J. M. Gere, *Theory of Elastic Stability*, 2nd ed (McGraw-Hill, New York, 1961), p. 363.
- ²⁵D. A. Hill, X. Huang, G. Bazan, and G. H. Bernstein, *J. Appl. Phys.* **72**, 4088 (1992).
- ²⁶J. S. Greeneich, *J. Vac. Sci. Technol.* **16**, 1749 (1979).
- ²⁷L. Reimer, in *Scanning Electron Microscopy* (Springer, New York, 1985), p. 137.

Electron beam lithography over large scan fields

Greg Bazán¹ and Gary H. Bernstein

Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

(Received 30 September 1992; accepted 18 January 1993)

We have developed an advanced, research-oriented electron beam lithography system using a three channel arbitrary wave form generator and a scanning electron microscope. The system is capable of exposing patterns with features less than 100 nm over scan fields greater than 1.0 mm, as well as features less than 20 nm in smaller scan fields. Other system features include capabilities for registration and for generating exposure matrices for evaluating new electron beam resists and optimizing exposure parameters. We report the details of the system and show results made possible by our system.

I. INTRODUCTION

For more than a decade, intense effort has been put towards the construction of dimensionally confined systems for the study of quantum effects in mesoscopic structures.¹ To this end, simple devices have been constructed using processing methods similar to those used to fabricate GaAs metal-semiconductor field effect transistors (MESFETs). To study quantum effects, minimum feature sizes within these structures must be less than 1 μm , and often less than 100 nm.²⁻⁴ Several approaches for fabricating structures with features less than 50 nm have been implemented,⁵⁻⁷ but none combines the versatility and resolution capabilities of electron beam lithography (EBL) systems. X-ray lithography systems are capable of such resolution but are not readily available. Some commercial EBL systems have the resolution necessary to produce the required sizes yet are not readily available due to their cost. Using scanning electron microscope (SEM)-based EBL systems can provide a low-cost, versatile, research exposure tool whose only disadvantage is throughput. Other cost effective methods include the modification of scanning tunneling microscopes (STM),⁸ but this technology currently is limited by the overhead exposure time and the small exposure area.

Several EBL system designs have been implemented and described by others.⁹⁻²⁰ However, our work extends the utility of this technique into the regime of very large scan fields with nanometer resolution. Our motivation behind the development of this system was the ability to fabricate nanostructures and nanodevices incorporating all exposure steps (mesa, contacts, gates, bonding areas) into one tool as well as gate level exposure for quantum based circuit architectures.²¹ Also recent optical experiments require nanostructures in large scan fields due to the difficulty in focusing lasers onto small areas.²² We designed our system with three specific goals. The first was to build a system capable of exposing any arbitrary pattern with a resolution less than 100 nm over scan fields as large as 1.4 mm. The second was to allow the exposure tool to perform accurate pattern registration of 100 nm over scan fields as large as 1.4 mm. Finally, the system was designed so that cumula-

tive exposure parameters could be stored for future analysis and images printed and saved easily.

The most common method used to convert a SEM to an EBL system incorporates a computer fitted with two digital-to-analog converters (DACs) whose outputs are connected through amplifiers to the SEM scan coils. More sophisticated conversions involve computer controlled beam blanking and stage movement. Nabity and Wybourne⁹ give an especially thorough review of the design issues involved in a SEM-based EBL system. In all cases, the computer controls the beam position by the application of voltages via the DACs. Ideally, the beam position resolution of these DAC-based systems is strictly limited by the DAC resolution. Earlier designs used 12-bit DACs while most current systems use 16-bit DACs and, presumably, future systems will have even higher resolution. When 16-bit resolution is used, the beam can be positioned anywhere on a 65 536 by 65 536 point grid. The spatial resolution of this grid is determined by the SEM magnification. To move the beam, successive positions in the form of digital integers or pixels are input to the DAC at a predetermined rate calculated from the feature size, the exposure dose, and the beam current. This rate is calculated to include the exposure time at that location as well as any overhead time necessary to account for scan coil settling.⁹⁻¹¹

The minimum feature size of such systems is not only determined by the DAC spatial resolution, but also by the temporal resolution. For systems with poor temporal resolution, a low exposure dose can only be achieved by having fewer pixels in an exposure to result in the desired dose at a given magnification and beam current. In this case, if the pixel spacing is large compared to the beam diameter a discrete pixel exposure will result. Therefore, if faster pixel outputs are possible, pixels can be placed closer together on the grid resulting in smoother lines, more uniform exposure doses, and improved overall performance.

II. SYSTEM OVERVIEW

Control of the lithography system is achieved through an in-house software program²³ written specifically for electron beam lithography. The program we developed is capable of generating two types of files: layout exposure and alignment (registration). Each file must be assigned a

¹ Author to whom correspondence should be addressed.

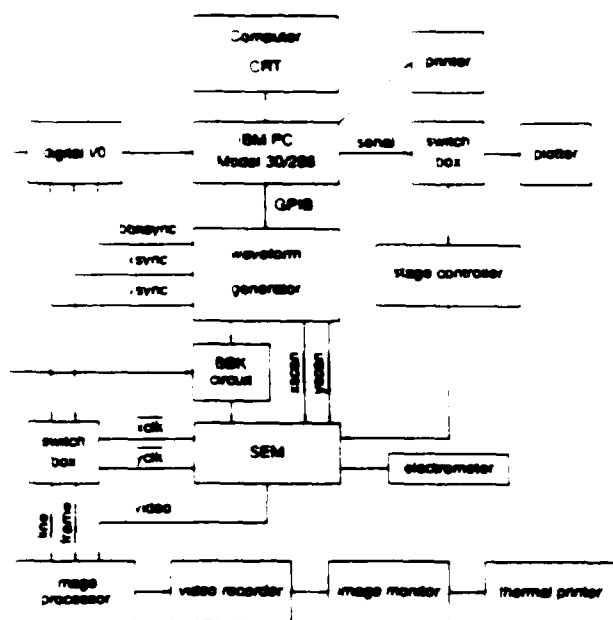


FIG. 1. Schematic diagram of electron beam lithography system.

magnification which is used to determine the scan field by the relation that a SEM magnification times the scan field equals a system constant. Our system constant has been set at 140 000 where the scan field is in units of microns. Having such a large system constant allows for the possibility to expose patterns over 7 mm. Both files allow for an exposure to occur only if the beam current and dose are specified. A layout file can consist of dot, line, and area exposures defined as a set of primitive shapes. Currently the set of shapes includes dot grids, angled lines, gratings, grids, rings, rectangles, parallelograms, and triangles. The registration process involves continuously scanning or imaging any features designated as the alignment marks. These features must be positioned at the corners of a rectangle so that adjustments can be made to reduce rotational, horizontal, and vertical overlay errors as well as to select the correct magnification. These marks are scanned simultaneously or individually through a window whose size can be varied to improve the alignment precision.

A complete schematic diagram of our lithography system is shown in Fig. 1. The lithography system is controlled by an IBM PC Model 30/286²⁴ equipped with a 32 megabyte hard drive, 4 megabytes of random access memory (RAM), a Hewlett-Packard general purpose interface bus (GPIB) interface card, and a Computer Boards, Inc. CIO-DIO24 digital input/output card. The GPIB interface controls the operation of the wave form generator and the CIO-DIO24 card monitors the channel synchronization outputs of the wave form generator and controls the synchronization of the beam blanker. Connected to the computer through the serial port is a switch box used to direct information to a plotter for layouts or to the SEM stage control system. The computer is also connected to a laser-printer to record SEM exposure parameters as well as the size and position of exposures within a layout. Our SEM is an Amray 1400 with a 50 kV acceleration potential, tung-

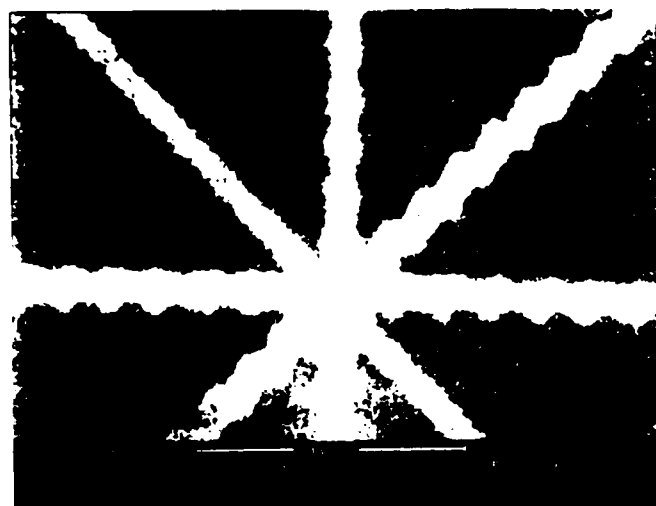
sten source, electrostatic beam blanker, and a Robinson backscattered electron detector. Such a large acceleration potential gives rise to a wide backscattered electron distribution in the resist, thus reducing proximity effects during the exposure of dense layouts. The Robinson backscattered electron detector is used during the registration process to image marks through the resist. Also connected to the SEM is a Keithley 614 electrometer used to measure beam current collected in a Faraday cup mounted to the sample holder.

The beam position and blanking are controlled by a Pragmatic Instruments Model 2201A high-definition arbitrary wave form generator (AWG). Current advances in wave form generation electronics^{26,27} make the use of a high resolution AWG a superior choice for controlling SEM-based EBL systems. The 2201A is a three channel wave form generator with each channel capable of generating a wave form with a maximum length of 65 535 samples and sampled at a selectable rate from 0.100 Hz to 2.000 MHz with four-digit resolution. Each wave form sample has 16-bit resolution within a selectable voltage range from 0.01 to 10.2 V with four-digit resolution. The selectable channel amplitude is extremely useful for SEMs whose scan amplifiers compensate for the size of micrographs. By scaling the output of the 2201A accordingly, the SEM scan amplifier ratio can be balanced resulting in a square scan field. Each channel also has a synchronization output which is used to monitor the end of wave form generation, to provide the pulse signals needed for registration, and to control the beam blanking. The three wave form outputs are connected to the SEM with channel 1 controlling beam blanking, and channel 2 and 3 functioning as analog outputs to control the horizontal and vertical beam positions. We refer to channel signal outputs as BBK, XSCAN, and YSCAN, respectively, and channel synchronization outputs are referred to as BBKSYNC, XSYNC, and YSYNC, respectively.

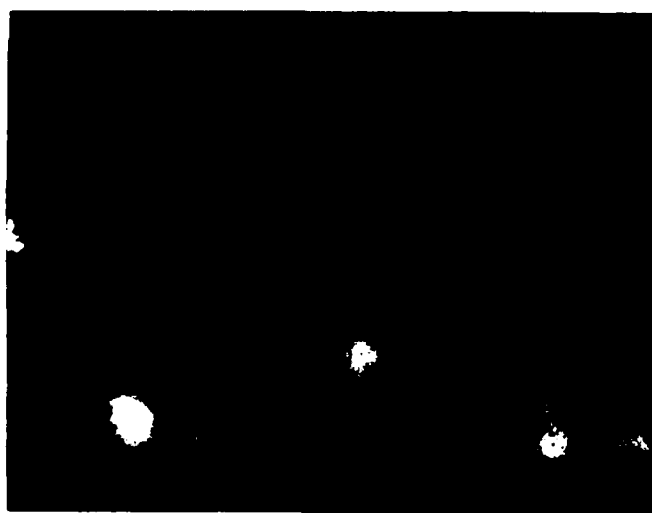
The Quantel Crystal Digital Image Processor is used to capture video information from the SEM. To capture images, the image processor requires video input as well as line and frame pulse signals. These pulse signals correspond to the end of a horizontal scan and the end of a vertical scan, respectively. To capture an image, the processor synchronizes its capture rate to that of the line and frame pulses. When the lithography system is used in an imaging mode, the image processor receives line and frame pulses from the SEM and the scanned image is digitized. When the system is used in a registration mode, the processor receives line and frame pulses from the 2201A (XSYNC and YSYNC) so that only the scanned areas over the registration marks are digitized.

III. LAYOUT EXPOSURE

We have taken great care in designing an EBL system capable of large scan field exposures with nanometer resolution. For us, the primary issue for large scan field exposures is the signal-to-noise (S/N) ratio in the signals used to control the SEM scan coils. Noise in the generated signal can result in a deflection of the beam which depends on



(a)



(b)

Fig. 2 Improvements made in noise reduction for an exposure completed in a 1.4 mm scan field. (a) Pattern processed before noise reduction. (b) Pattern after improvements were made to the system.

the gain of the scan amplifiers. In order to complete patterns in large scan fields where the gain of the scan amplifiers is large, the noise must be less than the minimum feature size desired. For example, in a scan field of 0.7 mm, for a 30 nm feature to be exposed, the S/N ratio of the signal generator must be greater than 87 dB. We carefully analyzed the scan coil signals at various points in the lithography system using a Hewlett-Packard 35660A dynamic signal analyzer and were able to improve the S/N ratio measured at the SEM scan coils from 78 to 100 dB using proper shielding and grounding techniques.^{28,29} A 100 dB S/N ratio represents a peak-to-peak noise signal equivalent to only 14 nm in a 1.4 mm scan field.

Every stage of the system from the computer to the scan coils was modified. Substantial modifications were required to eliminate signals emanating from the filament current power supply. By isolating the supply from the scan coil electronics power supply, the S/N ratio improved by 10 dB. The primary source of noise was identified as a ground

loop problem resulting from the GPIB connection between the 2201A and the computer. Two commercial GPIB optical isolation units were tested with no improvement noted. Therefore, a ground network was designed and implemented resulting in a 20 dB improvement in the S/N ratio. Figure 2 shows a line pattern exposure before and after improvements were made. For large scan fields, the S/N ratio becomes much more of a factor as the noise in the scan coil signals directly correlates to an observable physical displacement of the beam as shown in Fig. 2(a). For small scan fields, the S/N ratio becomes much less of a limiting factor.

Layout and pattern exposure require the user to specify a beam current and magnification for the SEM, and exposure doses for the primitive shapes in the layout. The layout is exposed by programming the 2201A and CIO-DIO24 for each shape. The shape exposure order can be changed before pattern exposure. First, the 2201A is programmed with the necessary signals calculated to give the correct dose, for a specific magnification and beam current. Because of the flexibility of the 2201A, the shape algorithms are written so that the smallest time per pixel is calculated, allowing the highest resolution wave forms to be generated. Before a shape is exposed, a channel is selected to trigger the output of the other channels. The synchronization signal of this channel is then monitored by the CIO-DIO24 to detect the end of shape exposure. The AWG is then set to be triggered through the GPIB once the BBK circuit is programmed. The BBK circuit consists of fast logic gates controlled by the AWG synchronization outputs and the CIO-DIO24. The CIO-DIO24 selects which synchronization signals are to be used to control the SEM beam blanker. Also, the CIO-DIO24 is programmed to monitor the synchronization channel selected to signal the end of shape exposure. By continuously polling the CIO-DIO24 during shape exposure, the completion of an

TABLE I Commands used to configure 2201A for horizontal rectangle exposure

Commands	BBK	XSCAN	YSCAN
Amplitude	2.5	10.0	10.0
Offset	2.5	0	0
Mode	burst	burst	trig
Start address	0	0	0
Stop address	xsamples + xretrace - 1	xsamples + xretrace - 1	ysamples + 1
Burst	ysamples	ysamples	
Clock sel	twomeg	twomeg	synth
Filter		on	on
Degitch		off	off
Sync start			1
Sync length			ysamples
Output switch	on	off/on	off/on
Trigger sel	chan3	chan3	man
Trig1 start			1
Trig1 length			1
Trig2 start			1
Trig2 length			1

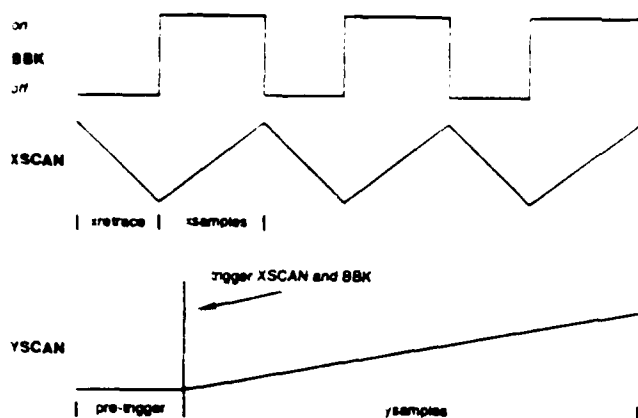


FIG. 3. Example of the 2201A wave form output when exposing a horizontal rectangle. Shown here is an example of three passes across the rectangle. YSCAN triggers BBK and XSCAN. Notice the beam is off during the retrace of XSCAN. With this, each sample in YSCAN corresponds to one XSCAN across the rectangle. The time scale for BBK and XSCAN is kept constant at $0.5 \mu\text{s}$ and the time scale of YSCAN is equal to the total time for one cycle of XSCAN ($x_{\text{retrace}} + x_{\text{samples}}$). To ensure channel synchronization, YSCAN wave form output begins on its SECOND wave form sample.

exposure is detected when 0.0 V is read from the triggering channel synchronization output.

The commands used to program the 2201A to expose a horizontal rectangle are shown in Table I. YSCAN is selected as the triggering channel for BBK and XSCAN signals. Figure 3 shows an example of the wave forms programmed to expose a horizontal rectangle from the instructions in Table I. The rectangle is exposed when YSCAN is triggered by the computer. As YSCAN is generated, when the second wave form sample of YSCAN is reached, BBK and XSCAN are triggered. This sequence guarantees that BBK and XSCAN are synchronized to YSCAN. The end of the exposure is detected by polling YSYNC through the CIO-DIO24. When YSCAN wave form output begins, YSYNC changes from 0.0 to 5.0 V. At the last wave form sample of YSCAN, YSYNC changes to a 0.0 V signaling the end of the shape exposure and the 2201A is programmed for the next shape exposure.

Use of the 2201A allows us to minimize the total overhead time for an exposure, as well as maximize control of

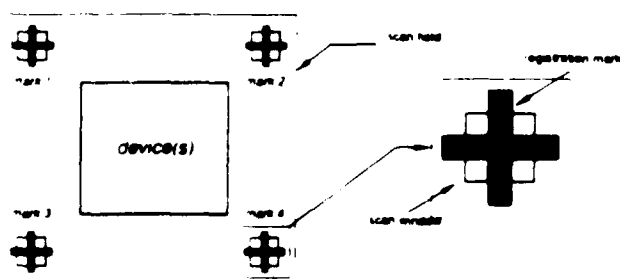
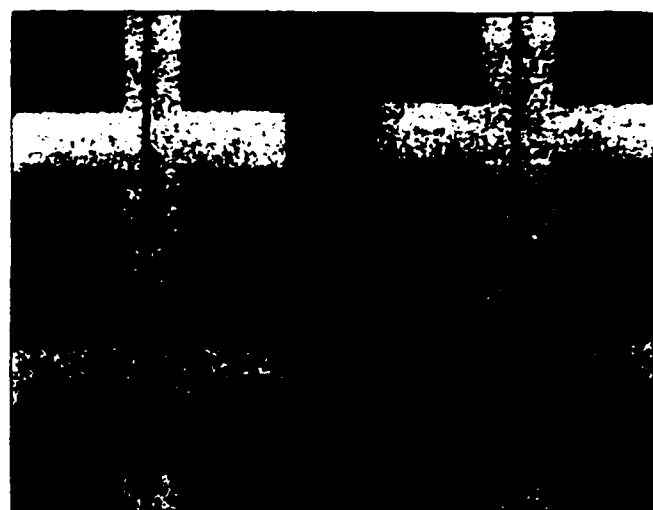
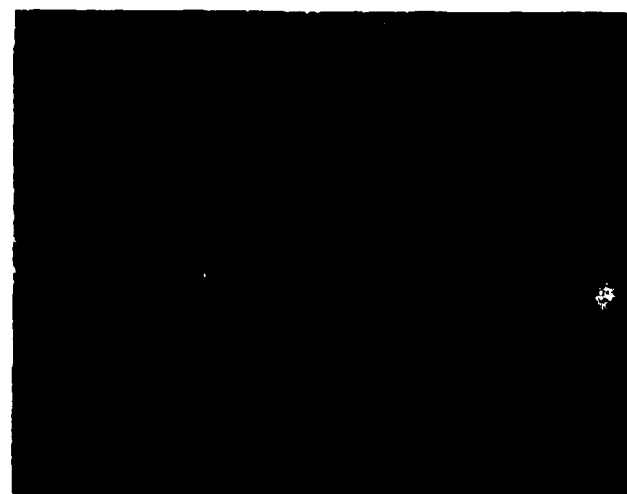


FIG. 4. Registration mark layout with definitions. Marks are numbered in a clockwise pattern. A scan of all marks is completed by scanning marks 1, 2, 3, and 4 in order. Scanning of individual marks is selected from a menu within the program. The size of the marks is enlarged for clarity.



(a)



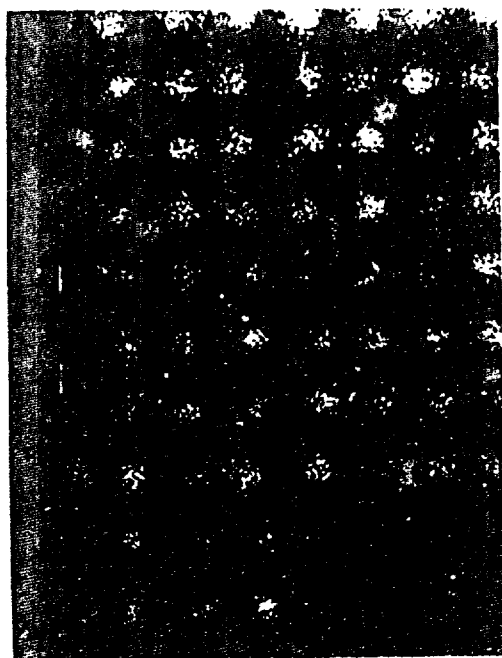
(b)

FIG. 5. Processed backscattered electron images of alignment marks for (a) four mark and (b) single mark scans.

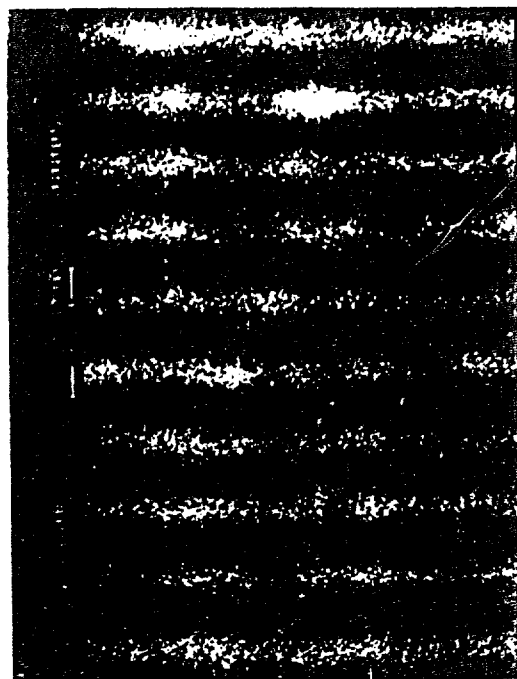
exposure dose. Because SEM parameters such as focus and astigmatism can drift, large overhead times can result in poor exposure conditions for complex, lengthy patterns. We have optimized the software for all line types to take advantage of this property. When a wave form is calculated, the 2201A is instructed to load its memory with the wave form samples. Next, a sampling rate is selected for each memory location and the wave form is then triggered for output. For the next exposure, instead of loading a new wave form, the start and stop pointers in channel memory are moved to correspond to the next exposure and triggered. With the resolution and range of sampling rates, the exposure dose can be varied by 0.1% from one exposure to another. Such resolution is needed when attempting to characterize new resists and developer solutions as well as to optimize exposure parameters.

IV. REGISTRATION

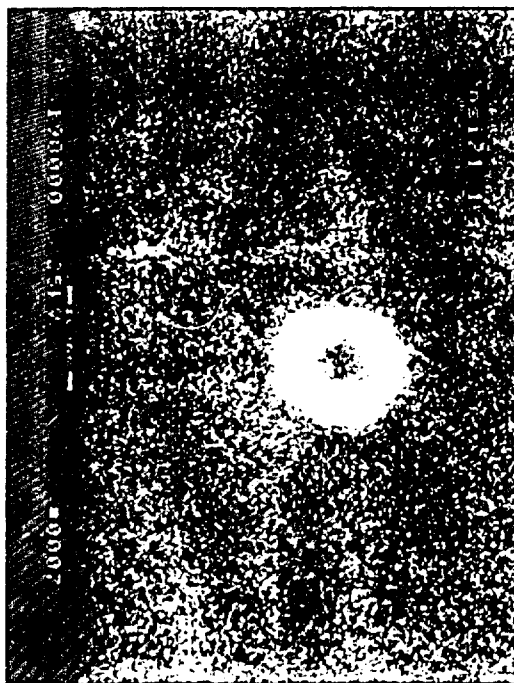
We have developed an accurate technique for achieving pattern registration, relying on the 2201A and the image



(a)



(b)

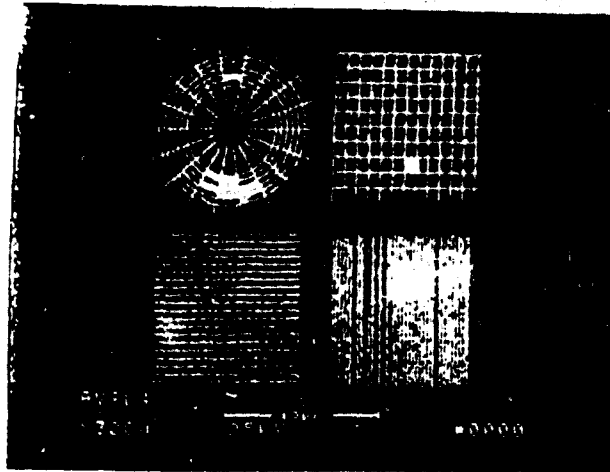


(c)

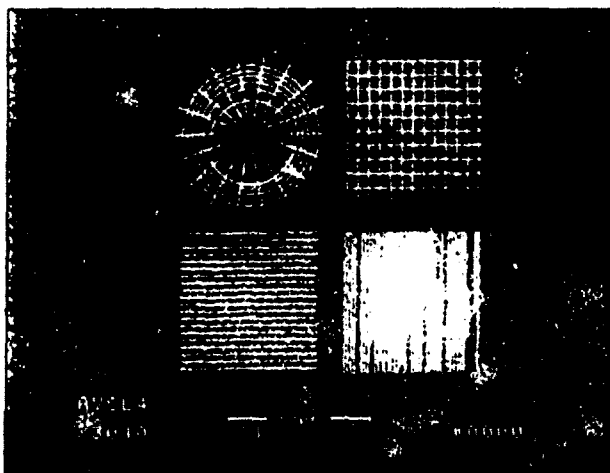


(d)

Fig. 6. Miscellaneous patterns demonstrating flexibility and resolution of the lithography system: (a) Au dots on 40-50 nm grid and (b) Au grating with lines on 55 nm centers exposed in the center of a 240 nm scan field. (c) Au ring with inner diameter less than 60 nm with a width of 18 nm. (d) Au line structure with linewidths less than 100 nm demonstrating arbitrary pattern exposure capability.



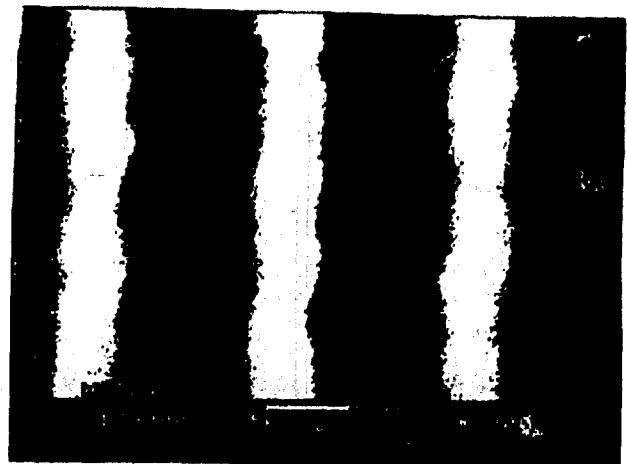
(a)



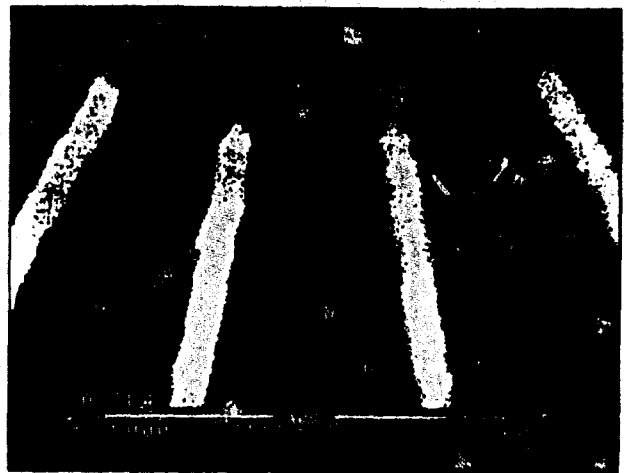
(b)

FIG. 7. Micrographs of line patterns exposed (a) in the center and (b) at the top edge of a 1.0 mm scan field.

processor. The 2201A is used to control the acquisition rate of the processor such that only the information scanned from the alignment marks is digitized. The synchronization signals of the 2201A are configured to emulate line and frame pulse signals with one frame consisting of 480 lines of video information. Shown in Fig. 4, our technique involves scanning one mark or scanning four marks simultaneously. The four-mark scan can be used to correct rotational errors while single mark scans are used to correct magnification errors as well as horizontal and vertical displacements. Once a mark has been scanned, the video information is collected by the processor and displayed. Examples of single and multiple mark scans are provided in Fig. 5. The superimposed dark areas in the scanned images are achieved through beam blanking and serve as guides during alignment. To improve registration accuracy, the size of the scan windows can be scaled, resulting in a magnified view of the registration mark



(a)



(b)

FIG. 8. Magnified view of lines in Fig. 7(b). (a) Vertical grating and (b) spoke pattern demonstrating features under 100 nm in a 1.0 mm scan field.

V. RESULTS

We have tested our system in a variety of scan fields, exposing arbitrary patterns with many exposure doses. Studies investigating the properties of poly(methylmethacrylate) (PMMA)^{30,31} and charges in oxides³² have been performed using this system. For this work, exposures were performed at 50 kV on thick silicon substrates with 100-nm-thick PMMA of molecular weight 950 000 a.m.u. baked for 10 min at 170 °C using a hot plate. Development was performed at 21 °C for 15 s in a 1:3 mixture of methyl isobutyl ketone: isopropanol with 1% methyl ethyl ketone.³³ Metal films of Au and Ti/Au were evaporated at normal incidence in an electron beam evaporator. Lift-off was performed by soaking in acetone for 2 to 3 min.

At conventional scan fields around 250 μm^2 , dots approximately 13 nm in diameter were fabricated.³⁴ Some of our results demonstrating the flexibility of the system, as well as resolution, are shown in Fig. 6. Figure 7 shows the same line pattern exposure at the center and top edge of a

1.0 mm scan field. Figure 8 shows magnified views of the vertical grating and spoke pattern at the edge of the 1.0 mm scan field demonstrating high resolution features in large scan fields. In Fig. 7(b), we see that the innermost ring is displaced in the vertical direction, which is an example of the effects of scan coil settling. The inner most ring is drawn first in this pattern. Once the scan coils have settled, the remaining pattern is exposed without any noticeable errors. Such errors are corrected by choice of pat-

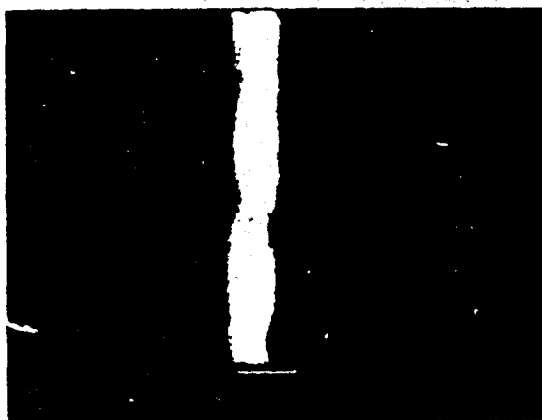
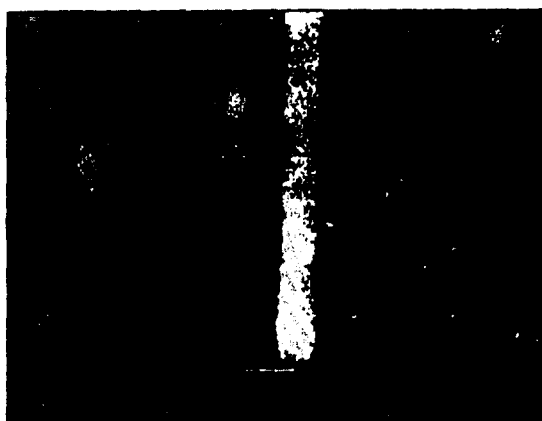
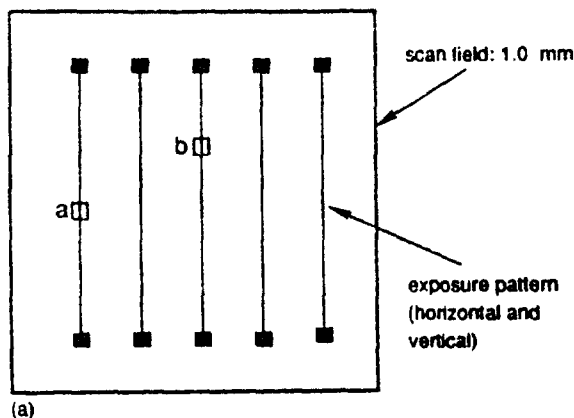


FIG. 9. Exposure pattern used to investigate the line width variation across a 1.0 mm scan field. The pattern was exposed both horizontally and vertically. Over 40 micrographs were taken of fabricated lines. Micrographs (a) and (b) correspond to location of boxes in figure drawn.

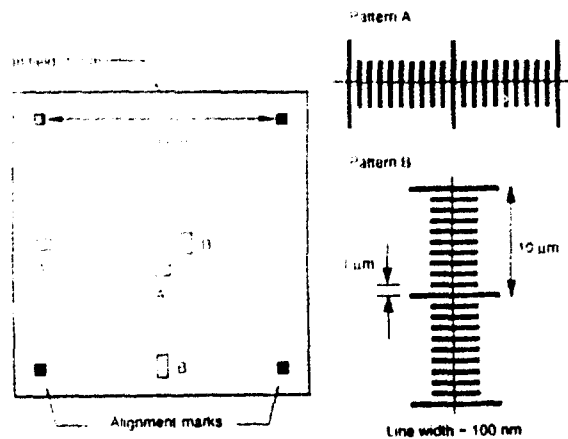


FIG. 10. Exposure pattern used to determine registration accuracy. The dark lines of patterns A and B were fabricated first. Having aligned the sample, the mirror image of patterns A and B was then fabricated and the registration accuracy determined.

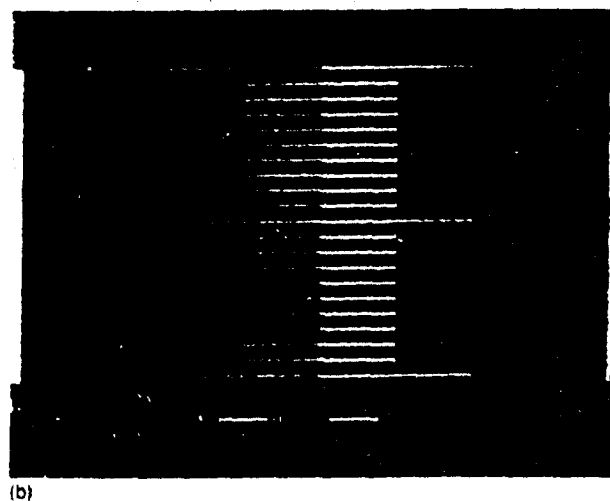
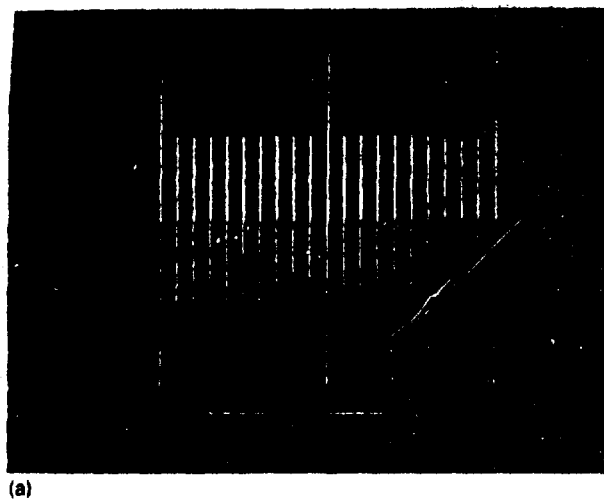


FIG. 11. Micrographs of alignment results for pattern A (a) and pattern B (b) of Fig. 10 located in the center of the scan field.

tern order and addition of delays into the 2201A programming sequence. To test line width variation over a 1.0 mm scan field, the pattern in Fig. 9 was exposed. The line variation was estimated to be 5 nm for 65-nm-wide lines. A more accurate investigation would require micrographs with more detail which, at this time, are currently not available. To test the registration accuracy, the double exposure pattern in Fig. 10 was used with our best results shown in Fig. 11. We believe we have achieved an overlay accuracy of less than 100 nm.

VI. SUMMARY

We have shown that using an AWG to control a SEM for EBL can provide a versatile nanolithographic exposure tool capable of large scan field exposures. With the flexibility of the 2201A, we are able to precisely control the exposure dose, accurately position the electron beam, and maintain beam position as well as control the image processor to facilitate device registration. The 16-bit resolution in voltage and four-digit resolution in sampling rates allows for optimal control of exposure dose. The variable output of the wave form generator also simplifies the adjustment of scan field aspect ratios. We have shown that we are capable of exposing patterns in large scan fields with nanolithographic features. Having the flexibility to pattern the fine structures as well as other device features in one system reduces the overhead associated with device design and fabrication.

ACKNOWLEDGMENTS

This work supported in part by IBM, NASA, Air Force Office of Scientific Research, Office of Naval Research, and the University of Notre Dame. We thank Xiaokang Huang for the micrographs of exposures completed in small scan fields. We especially thank Meinrad Steiner at Pragmatic Instruments, Inc. for his GPIB programming tips.

¹T. K. Gaylord, E. N. Glytsis, G. N. Henderson, K. P. Martin, D. B. Walker, D. W. Wilson, K. F. Brennan, *Proc. IEEE* **79**, 1159 (1991).

²J. M. Ryan, J. Han, A. M. Krizan, and D. K. Ferry, *Proc. SPIE* **1284**, 39 (1990).

³G. Bernstein and D. K. Ferry, *IEEE Trans. Electron Devices* **ED-35**, 887 (1988).

⁴G. Bernstein and D. K. Ferry, *Superlatt. Microstructures* **2**, 373 (1986).

⁵N. Giordano, *Phys. Rev. B* **22**, 5635 (1980).

⁶E. H. Anderson, C. M. Horwitz, and H. I. Smith, *Appl. Phys. Lett.* **49**, 874 (1983).

⁷D. C. Flanders, *Appl. Phys. Lett.* **36**, 93 (1980).

⁸M. A. McCord and R. F. W. Pease, *J. Vac. Sci. Technol. B* **6**, 293 (1988).

⁹J. C. Nability and M. N. Wybourne, *Rev. Sci. Instrum.* **60**, 27 (1988).

¹⁰S. Thoms, S. P. Beaumont, and C. D. W. Wilkinson, *J. Vac. Sci. Technol. B* **7**, 1823 (1989).

¹¹C. C. Andrews and W. P. Kirk, *J. Vac. Sci. Technol. B* **8**, 1858 (1990).

¹²A. J. Speth, A. D. Wilson, A. Lern, and T. H. P. Chang, *J. Vac. Sci. Technol.* **12**, 1235 (1975).

¹³J. Bellessa, F. Carcenac, A. Izrael, H. Launois, and D. Mailly, *Micro. Elec. Eng.* **6**, 175 (1987).

¹⁴S. Mackie and S. P. Beaumont, *Solid State Technol.* **28**, 117 (1985).

¹⁵B. L. McIntyre and C. L. Dennis, *Appl. Opt.* **27**, 196 (1988).

¹⁶B. McIntyre, R. Perman, and B. Gold, *Res. Dev.* **31**, 44 (1989).

¹⁷C. F. Cook, T. F. AuCoin, and G. J. Iafrate, *Solid State Technol.* **28**, 125 (1985).

¹⁸C. P. Umbach, A. N. Broers, R. H. Koch, C. G. Willson, and R. B. Laibowitz, *IBM J. Res. Dev.* **32**, 454 (1988).

¹⁹A. Murray, M. Issacson, I. Adesida, and B. Whitehead, *J. Vac. Sci. Technol. B* **1**, 1091 (1983).

²⁰A. N. Broers, J. M. E. Harper, and W. W. Molzen, *Appl. Phys. Lett.* **35**, 392 (1978).

²¹C. S. Lent, P. D. Tougaw, W. Porod, and G. H. Bernstein, *Nanotechnology* (to be published).

²²B. McCombe (private communication).

²³Turbo Pascal version 5.5, Borland International.

²⁴A high speed computer is not necessary because the timing is performed by the high speed electronics of the AWG.

²⁵X. Huang, G. Bazán, G. H. Bernstein, and D. A. Hill, *J. Vac. Sci. Technol. B* (to be published).

²⁶A. Pini, *Eval. Eng.* **30**, 26 (1991).

²⁷J. Sauer, *Eval. Engr.* **31**, 26 (1992).

²⁸R. Morrison and W. H. Lewis, *Grounding and Shielding in Facilities* (Wiley, New York, 1990).

²⁹R. Morrison, *Noise and Other Interfering Signals* (Wiley, New York, 1992).

³⁰D. A. Hill, X. Huang, G. Bazán, and G. H. Bernstein, *J. Appl. Phys.* **72**, 4088 (1992).

³¹X. Huang, G. Bazán, D. A. Hill, and G. H. Bernstein, *J. Electrochem. Soc.* **139**, 2952 (1992).

³²G. H. Bernstein, S. W. Polchlopek, R. Kamath, and W. Porod, *Scanning* **14**, 345 (1992).

³³G. H. Bernstein, D. A. Hill, and W. P. Liu, *J. Appl. Phys.* **71**, 4066 (1992).

³⁴As imaged with a field emission SEM courtesy of Topcon Technologies, Inc.

New technique for computation and challenges for electron-beam lithography

Xiaokang Huang, Greg Bazán, and Gary H. Bernstein

Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556

(Received 15 June 1993; accepted 20 August 1993)

In this article, the basic concepts of our recently proposed computing architecture based on Coulomb coupling of nanofabricated structures, called quantum cellular automata (QCA) are reviewed and fabrication issues critical to the new technology are discussed. The QCA fabrication will require an extremely high level of lithographic control. To this end, the proximity effects in making very high density patterns with poly(methylmethacrylate) (PMMA) and electron-beam lithography have been experimentally investigated. A triple Gaussian model was used to simulate the experimental data. By using a 50 keV electron beam, sub-40 nm pitch gratings, double lines, and dot grids were successfully fabricated on Si and SiO₂/Si bulk wafers with single-level PMMA and lift-off.

I. INTRODUCTION

High density patterns are very important for such structures as quantum devices and high speed photodetectors,¹ as well as studies of basic quantum phenomena, such as current drag.² Recently, we proposed a new paradigm for computing with cellular automata (CA) composed of arrays of quantum dots arranged into separate cells of several dots.^{3,4} We call this new architecture "quantum cellular automata" (QCA). Our calculations show that the dots, which are fundamental to the new architecture (discussed below), must be as small as 10–15 nm in diameter on a pitch of 15–30 nm with intracell variations of at most 5%–10% over a few micron area. This implies that electron-beam lithography (EBL) to be used in the fabrication will be put to a serious test. Although Allee *et al.*^{5,6} have directly patterned 15 nm gratings on SiO₂/Si with EB irradiation, lift-off with poly(methylmethacrylate) (PMMA) is still one of the most common techniques used in pattern transfer. The latter has so far produced results in the 40–50 nm pitch range.¹ In this size regime, pitch is limited by the strength of PMMA and proximity effects during exposure.

Proximity effects result from the distributions of injected and scattered electrons in the resist. Because of the distributions, the resist which is not directly addressed by the primary EB will also be exposed. The electron distributions in resist have been discussed in many papers,^{7–9} and a triple Gaussian model^{8–10} was used in modeling the distributions. The triple Gaussian model is

$$f(r) = \frac{1}{\pi(1+\eta+\eta')} \left[\frac{1}{\alpha^2} \exp\left(-\frac{r^2}{\alpha^2}\right) + \frac{\eta}{\beta^2} \exp\left(-\frac{r^2}{\beta^2}\right) + \frac{\eta'}{\gamma^2} \exp\left(-\frac{r^2}{\gamma^2}\right) \right]. \quad (1)$$

The first term includes the distributions of the primary electron beam and forward scattered electrons. The width of the term α ranges from a few nm to tens of nm, which depends on the beam size and energy, and the resist. The second term only includes the backscattered electrons from the substrate. The width of the distribution of backscat-

tered electrons β is a few microns and depends on beam energies and types of substrates. The higher the beam energy, the wider the distribution.¹¹ The third term, the "broad range electrons,"¹⁰ describes the behavior of all other electrons that are not included in the first two terms.^{8,9} Its width γ is a few hundreds of nm. Here, η and η' are the ratios of the exposures of the β and γ term to the forward exposure, respectively.

Because of their distributions, different electrons result in proximity effects in different ranges. Although there are many papers that have investigated the proximity effects in EBL^{8,9,12} and many methods for computer-aided proximity effect corrections during EBL have been developed,^{13,14} the feature sizes and spacing in most of the published work are in the half- or quarter-micron regime, which is important in ultra-large scale integration (ULSI) circuit fabrication. Since large throughput is required in manufacturing, low beam energy (20 keV) and large currents are used in industrial EBL. In that case, backscattered electrons can dominate the proximity effects. However, for making very high density patterns that approach the spatial density limit of EBL, high beam energy (50 keV or higher), and small beam size (smaller than 10 nm) are usually used. In applications of nanolithography for quantum devices, pattern sizes are usually very small compared to those of ULSI. Therefore, the proximity effects from forward scattered electrons and secondary electrons become more pronounced.¹⁰

In this article, we discuss the basic properties and requirements of the QCA architecture. Toward the achievement of the required size scales, we report an experimental investigation of proximity effects for lines with pitch down to 50 nm. A triple Gaussian model is used in the theoretical simulation of the experimental results. The purpose of discussing the proximity effects is not directly to make proximity corrections in processing, but to understand and overcome the difficulties in making high density patterns. A spatial density limit in gratings caused by the proximity effects in PMMA is also determined. We studied the proximity effects in gratings rather than in arrays of dots (which will be more typical of the QCA work to be done)

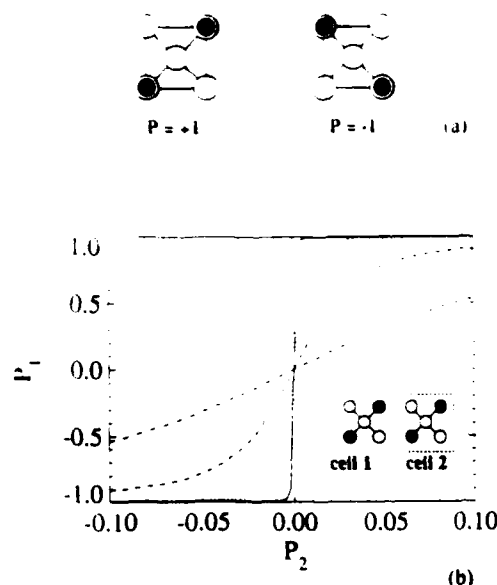


FIG. 1. (a) The quantum cell consisting of five quantum dots which are occupied by two electrons. The mutual Coulombic repulsion between the electrons results in bistability between the $P = +1$ and $P = -1$ states. (b) The cell-cell response function for various values of the dot-to-dot coupling energy t . The induced cell polarization P_1 is plotted as a function of the neighboring cell polarization P_2 . The results are shown for values of $t = -0.2$ (full curve), -0.3 (dotted curve), -0.5 (dashed curve), and -0.7 (dot-dashed curve) meV. Note that the response is shown only for P_2 in the range $[-0.1, +0.1]$.

since gratings are more susceptible to proximity effects due to larger integration of dose, and the results would be more easily interpretable. The results discussed below can be applied to structures which require spacing in the sub-50 nm range. Toward this goal, 38 nm pitch gratings, 36 nm pitch double lines, and 37 nm pitch dot grid on Si and SiO_2/Si bulk substrates were successfully made with 50 keV EBL combined with the lift-off technique.

II. QUANTUM CELLULAR AUTOMATA (QCA)

QCA consist of separate cells containing several quantum dots per cell, as shown in Fig. 1(a). We have proposed a specific realization of these ideas using two-electron cells composed of quantum dots, which we will later show is within the reach of current fabrication technology. We define polarization as an average of the two-electron wave function as defined in Refs. 3 and 4. If the charge is aligned along one axis as shown in Fig. 1(a), the polarization is "+1" (encoding the binary value "1") and when distributed along the other axis the polarization is "-1" (encoding the binary value "0"). The mutual Coulomb repulsion of the two electrons causes the charge density in the cell to be very highly polarized (aligned) along one of the two cell axes, suggestive of a two-state CA. Figure 1(b) shows the polarization of a two-electron cell due to the polarization of an adjoining cell for various values of the dot-to-dot coupling energy, t (which depends on the intracell barrier heights and intracell dot spacing). The polarization of one cell induces a polarization in a neighboring cell through the Coulomb interaction in a very nonlinear fashion.

Electrons within the cells are coupled quantum mechanically by tunneling between sites, but coupling between cells is strictly by Coulomb interactions. The cells are arranged in arrays which interact to perform computational functions. Computing in the new paradigm is edge driven. Input, output, and power are delivered to the edge of the array only; no direct flow of information or energy to internal cells is required. The architecture is so designed that the ground state configuration of the array, subject to boundary conditions determined by the inputs, yields the computational result. We refer to computations performed in this manner as "computing in the ground state."

Device modeling has been performed by solving the two-electron Schrödinger equation directly for each cell. Arrays of cells were modeled using a self-consistent Hartree technique based on the two-electron problem in each cell. We have shown^{3,4} that useful computing structures can be built from a set of logical primitives composed of quantum cells. This set includes wires, wire crossings, inverters, and a flexible three-input structure. The three-input device can be configured as an AND gate, an OR gate, or a majority logic unit. The reader is referred to the references for a complete description of these logic devices.

The precise nature of the cell fabrication is critical to the success of the CA. The QCA cell arrays are forgiving of variations of dot size and spacing when variations occur only between adjoining cells, but not within a cell. Figure 1(b) shows that the resulting polarization of a cell is very large when subjected to only a small polarization of the input cell. This implies that spacing between cells is not very critical since the Coulomb interactions affected by variations in distance are equivalent to variations of polarization. However, if a cell contains dots of various sizes, then it is prebiased toward one or the other state. Although this can be very useful in some cases,^{3,4} if uncontrolled, the state of the cell would be very difficult to control accurately. The remainder of this article is devoted to an experimental investigation of the role of proximity effects in obtaining very dense patterns such as those likely to be used in QCA fabrication.

III. EXPERIMENTAL PROCEDURE

Single layers of 950K PMMA 50–70 nm thick were spun on bulk Si and SiO_2/Si substrates. The thickness of the SiO_2 in all cases was 200 nm. Patterns were written at 50 keV with an Amray 1400 scanning electron microscope employing a W cathode, customized as an EBL system.¹⁵ After exposure, samples were developed with methylisobutylketone:isopropanol:methylethylketone (MIBK:IPA:MEK) (1:3:0.06) maintained at 26 °C. At this temperature, the contrast of the developer has been demonstrated to be larger than 10.¹⁶ Our samples were not rinsed with pure methanol, as is typically done following development.¹ We have found¹⁷ that methanol and other solutions weaken the strength of PMMA walls and limit the space between patterns, and that in fact this rinse step is not necessary. We found that replacing this step by drying in pure N_2 immediately after development is of benefit to the strength of the PMMA walls, and therefore helps to

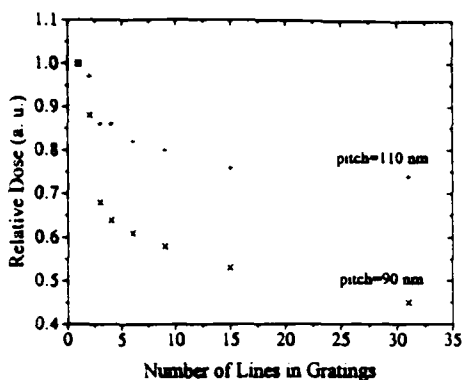


FIG. 2. Experimental results of the relative dose as a function of number of lines in gratings. +: 110 nm pitch gratings; x: 90 nm pitch gratings.

achieve high density patterns with lift-off. Au or Ti/Au was evaporated with an EB evaporator followed by lift-off with acetone soaking and shooting with a syringe.¹⁸

The relationship between resulting linewidth w and (single-pass) line dose D_{line} (defined in units of C/cm) is approximately linear at lower doses and saturates at higher doses,^{18,19} yielding a relationship for the average area dose D_{area} in the linear region as,

$$D_{\text{area}} = D_{\text{line}}/w (\text{C}/\text{cm}^2). \quad (2)$$

This implies that in all of our exposures at low doses and narrow lines, the average dose across the developed area of the line is constant regardless of the final linewidth.

IV. RESULTS AND DISCUSSION

As discussed above, the linewidth increases linearly with small exposure doses. If two lines are exposed near each other, each line will get extra exposure dose from the other line exposure because of proximity effects. In order to achieve the same linewidth, the average area dose defined by Eq. (2) should be smaller in two close lines than in a single line pattern. This means that the slopes in the relation between linewidth and exposure dose will be different in the two cases.¹⁸ We define relative dose as the ratio of the average area dose of multiline patterns to the average area dose of a single line pattern. Figure 2 shows the relative dose as a function of the number of lines in a pattern. As expected, relative dose decreases as more lines are added into the pattern because of interline exposure by the proximity effects. Also the relative dose decreases when lines get closer together. This can be seen more clearly in Fig. 3 which shows the relative dose as a function of the line pitch of gratings. The gratings in our study were $3 \mu\text{m} \times 3 \mu\text{m}$ separated by $2 \mu\text{m}$ spaces. Linewidths were measured at the center of gratings and ranged from 20 to 45 nm for different pitch gratings. This is different from ULSI in which equal space and linewidth are important,⁹ since this is not typically necessary for quantum device applications. In Fig. 3, the relative dose was defined as the ratio of the average area doses of gratings to those of the 150 nm pitch gratings. The dot-dashed line in Fig. 3 is a calculated result using Eq. (1).²⁰ In our calculation, we use

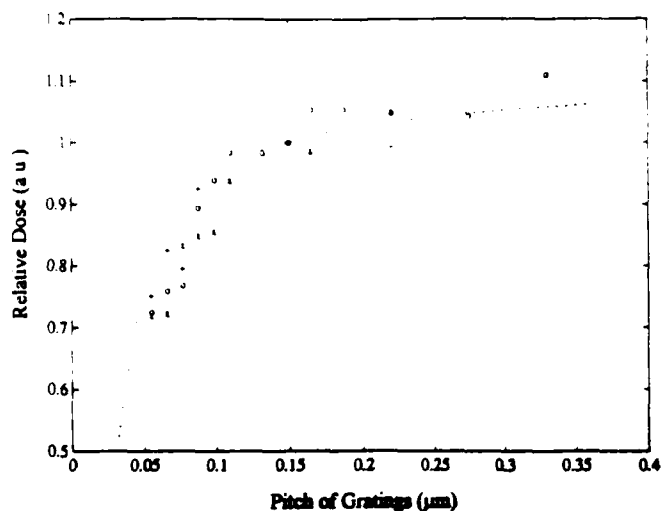


FIG. 3. The relation between the relative dose and the pitch of gratings. O, x, and + are separate sets of experimental results. The dot-dashed line is the simulation with the triple Gaussian model.

$\beta = 10 \mu\text{m}$ and $\eta = 0.8$ for 50 keV backscattered electrons in solid silicon substrates according to Ref. 11. Ideally, α should be determined by a convolution of the beam shape with the lateral scattering profile within the resist.²¹ Since this is complicated by knowledge of the exact shape of the beam and lateral scattering properties, we instead fit the experimental data.^{8,9,12} For the best fit, α , γ , and η' were chosen as $\alpha = 0.029 \mu\text{m}$, $\gamma = 0.4 \mu\text{m}$, and $\eta' = 0.35$. Although backscattered electrons from other gratings also contributed to the total dose, it is appropriate to neglect them in our case since $\eta/\beta^2 \ll 1/\alpha$ and η'/γ^2 , and the total area of all other gratings within a range of β away from the center was small. From Fig. 3, it can be seen that the relative dose decreases very quickly for line pitch smaller than 100 nm, while it is relatively flat for line pitch larger than 150 nm. Relative dose as a function of pitch of double lines also shows a similar relation.¹⁹

The different relationship between the relative dose and the pitch of gratings is due to the proximity effects from primary, forward, backscattered, and long-range electron distributions. This has been presented in Ref. 10. Figure 3 is important because it not only shows the significant proximity effects caused by forward scattered electrons and fast secondary electrons, but it is also very useful in predicting the dose needed to achieve certain pitch gratings with required linewidth in EBL.

By using Eq. (1), the minimum exposure received between lines (E_p) and the maximum exposure received by the lines (E_r) can be calculated. The larger the ratio E_r/E_p , the more difficult the task faced by the developer and the higher must be the contrast of the developer utilized in order to delineate the latent image. Figure 4 shows E_r/E_p as a function of the grating pitch. β , γ , η , and η' were the same as in the calculation in Fig. 3, while different widths of forward scattered electrons (α) were used in the calculation. The shift of the E_r/E_p curve to the left as α decreases means that higher spatial density can be achieved for smaller α . Also E_r/E_p changes with grating pitch and

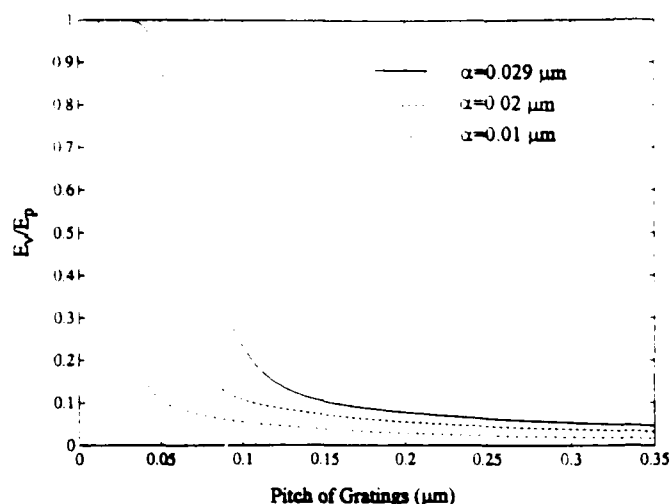
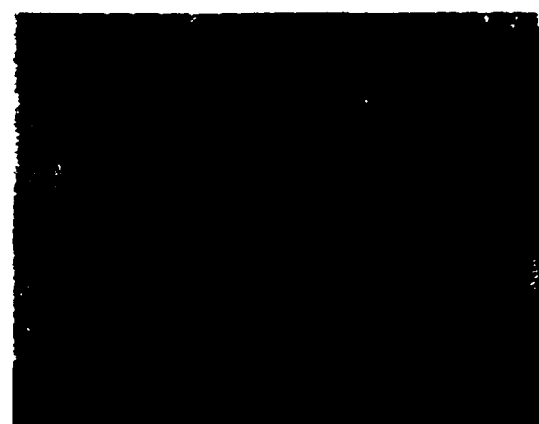


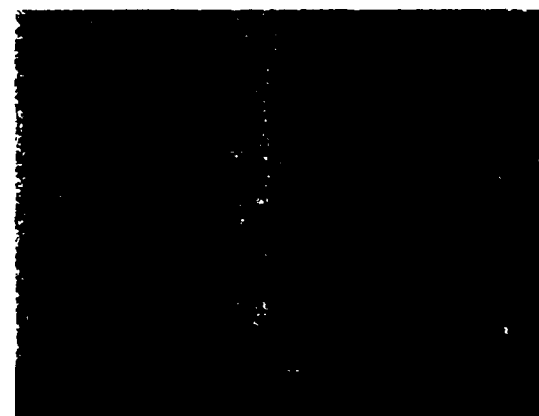
FIG. 4. E_v/E_p as a function of the pitch of gratings. E_v is the minimum exposure received between lines, and E_p is the maximum exposure received on the lines.

approaches unity when lines are very close. In that case, patterns are washed out. The role of the developer is to distinguish E_v and E_p to produce the required patterns. For a given developer contrast, there is a maximum E_v/E_p that can still be distinguished by the developer. E_v/E_p can be related to D_i/D_f , where D_i is the exposure dose at the onset of development and D_f is the exposure dose for complete development, and developer contrast γ^* is defined as $1/\log(D_i/D_f)$. Therefore, the minimum pitch that can be achieved with the developer can be obtained from Fig. 4. We used a very high contrast developer in our experiments. The contrast for our developer is greater than 10 at 26 °C,¹⁶ so it can distinguish E_p from E_v at $E_v/E_p \sim 0.8$ or even higher. Therefore, only 50–60 nm pitch gratings could be expected in our proximity effect experiments in which 4 pA beam current was used. However, with $\alpha = 10$ nm,⁷ 20 nm pitch gratings can be expected.

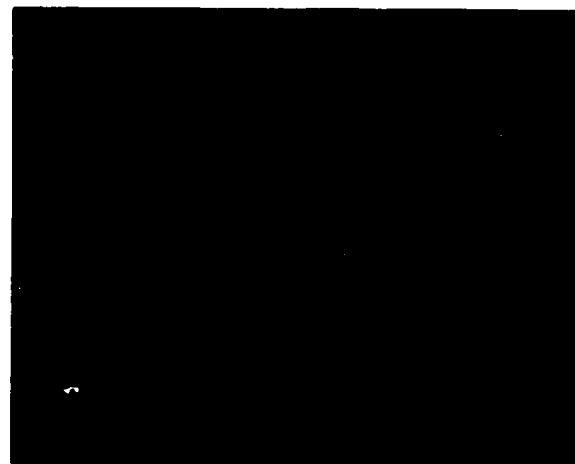
Guided by the discussion above, the beam current was decreased from 4 to 2.5 pA. Although this made focusing more difficult because of image noise, by carefully focusing of a narrower beam, we achieved a minimum grating pitch of 38 nm, as shown in Fig. 5(a). In order to get this result, Fig. 3 was used with Eq. (2) to determine the best line dose range needed in the exposure. Also, extreme care was taken during the development because any overdevelopment could cause the lines to wash out. These data points are not shown in Fig. 3 for two reasons. First, α was changed as the beam current decreased. Second, the linewidth of the 38 nm pitch grating could not be accurately measured because of the resolution of our scanning electron micrograph (SEM), so it was difficult to get the accurate area dose required in Fig. 3. The grating is made by Ti/Au (2 nm/15 nm) on top of SiO₂/Si by lift-off. 36 nm pitch double lines made by lift-off are shown in Fig. 5(b). Since there is less proximity effect in double lines than in gratings, as shown in Fig. 2, making dense double lines is easier than making dense gratings. However, besides proximity effects, the spatial density of EBL is also limited by



(a)



(b)



(c)

FIG. 5. (a) A Ti/Au (2 nm/15 nm) grating with 38 nm-pitch; (b) A lift-off Ti/Au (2 nm/15 nm) double lines with 36 nm center-to-center distance; both (a) and (b) are made by lift-off on SiO₂/Si substrates. (c) A gold (15 nm thick) dot grid on Si substrate with 37 nm pitch by lift-off

the minimum fundamentally achievable linewidth and space. The former (linewidth) is determined by beam size and fast secondary electrons (intraline proximity effect), and is ~ 10 nm.⁷ The latter (space) is dominated by the strength of PMMA,^{17,22} and is also ~ 10 nm for 45 nm thick resist. This means that the minimum achievable pitch of double lines is ~ 20 nm, i.e., the same as the minimum expected pitch of gratings limited by all proximity effects. A 37 nm pitch dot grid is shown in Fig. 5(c). It is 15 nm

thick Au on a Si substrate made by lift-off. This micrograph was obtained with a Topcon model ABT-150F field emission SEM. Since there is much less proximity effect in making dot grids, and the strength of PMMA is not as important in this case, 10 nm diam dots with 15 nm pitch dot grids can be expected.

V. SUMMARY AND CONCLUSIONS

We discussed a new paradigm in computing called quantum cellular automata and related the physical properties of the quantum structures to the limits of EBL. We investigated proximity effects in order to predict the ultimate size of the achievable dot arrays for QCA applications. We investigated properties of dense gratings and dot arrays with a minimum achieved double line pitch of 36 nm. We predict that dot arrays with a pitch of 15 nm should be achievable. Thus, the requirements for the fabrication of dense arrays of dots for applications to quantum cellular automata are predicted to be within the physical limits of lift-off technology using PMMA resist and EBL.

ACKNOWLEDGMENTS

The authors thank W. Porod, C. S. Lent, and P. D. Tougaw for helpful discussions. This work was supported by the National Science Foundation, the National Space Grant College and Fellowship Program, the Air Force Office of Scientific Research, and the Office of Naval Research. They also thank Topcon Technologies Inc. for providing field emission SEM micrographs.

- P. B. Fischer and S. Y. Chou, *Appl. Phys. Lett.* **62**, 2989 (1993).
- A. G. Rojo and G. D. Mahan, *Phys. Rev. Lett.* **68**, 2074 (1992).
- C. S. Lent, P. D. Tougaw, and W. Porod, *Appl. Phys. Lett.* **62**, 714 (1993).
- C. S. Lent, P. D. Tougaw, W. Porod, and G. H. Bernstein, *Nanotechnology* **4**, 49 (1993).
- D. R. Allee and A. N. Broers, *Appl. Phys. Lett.* **57**, 2271 (1990).
- D. R. Allee, C. P. Umbach, and A. N. Broers, *J. Vac. Sci. Technol. B* **9**, 2838 (1991).
- D. C. Joy, *Microelectron. Eng.* **1**, 103 (1983).
- R. J. Bojko and B. J. Hughes, *J. Vac. Sci. Technol. B* **8**, 1909 (1990).
- S. J. Wind, M. G. Rosenfield, G. Pepper, W. W. Molzen, and P. D. Gerber, *J. Vac. Sci. Technol. B* **7**, 1507 (1989); M. G. Rosenfield, S. J. Wind, W. W. Molzen, and P. D. Gerber, *Microelectron. Eng.* **11**, 617 (1990).
- X. Huang, G. H. Bernstein, G. Bazan, and D. A. Hill, *J. Vac. Sci. Technol. A* **11**, 1739 (1993).
- P. M. Mankiewich, L. D. Jackel, and R. E. Howard, *J. Vac. Sci. Technol. B* **3**, 174 (1985).
- S. A. Rishton and D. P. Kern, *J. Vac. Sci. Technol. B* **5**, 135 (1987).
- R. C. Frye, *J. Vac. Sci. Technol. B* **9**, 3054 (1991).
- S. Y. Lee, J. C. Jacob, C. M. Chen, J. A. McMillan, and N. C. McDonald, *J. Vac. Sci. Technol. B* **9**, 3048 (1991).
- G. Bazan and G. H. Bernstein, *J. Vac. Sci. Technol. A* **11**, 1745 (1993).
- G. H. Bernstein, D. A. Hill, and W.-P. Liu, *J. Appl. Phys.* **71**, 4066 (1992).
- D. A. Hill, X. Huang, G. Bazan, and G. H. Bernstein, *J. Appl. Phys.* **72**, 4088 (1992).
- S. Mackie and S. P. Beaumont, *Solid State Technol.* **117** (1985).
- X. Huang and G. H. Bernstein (unpublished).
- T. H. P. Chang, *J. Vac. Sci. Technol.* **12**, 1271 (1975).
- J. S. Greeneich and T. van Duzer, *IEEE Trans. Electron Device* **ED-21**, 286 (1974).
- X. Huang, G. Bazan, D. A. Hill, and G. H. Bernstein, *J. Electrochem. Soc.* **139**, 2852 (1992).